

*Registers in  
sociaalwetenschappelijk  
onderzoek*

*Mogelijkheden en valkuilen*

Onder redactie van:  
Bart F.M. Bakker  
Léander Kuijvenhoven

## Verklaring van tekens

.	= gegevens ontbreken
*	= voorlopig cijfer
**	= nader voorlopig cijfer
x	= geheim
–	= nihil
–	= (indien voorkomend tussen twee getallen) tot en met
0 (0,0)	= het getal is kleiner dan de helft van de gekozen eenheid
niets (blank)	= een cijfer kan op logische gronden niet voorkomen
2007-2008	= 2007 tot en met 2008
2007/2008	= het gemiddelde over de jaren 2007 tot en met 2008
2007/'08	= oogstjaar, boekjaar, schooljaar enz., beginnend in 2007 en eindigend in 2008
2004/'05-2007/'08	= oogstjaar, boekjaar enz., 2004/'05 tot en met 2007/'08

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

## Colofon

*Uitgever*  
Centraal Bureau voor de Statistiek  
Henri Faasdreef 312  
2492 JP Den Haag

*Prepress*  
Centraal Bureau voor de Statistiek  
Grafimedia

*Druk*  
OBT bv, Den Haag

*Omslag*  
TelDesign, Rotterdam

*Inlichtingen*  
Tel. (088) 570 70 70  
Fax (070) 337 59 94  
Via contact formulier:  
[www.cbs.nl/infoservice](http://www.cbs.nl/infoservice)

*Bestellingen*  
E-mail: [verkoop@cbs.nl](mailto:verkoop@cbs.nl)  
Fax (045) 570 62 68

*Internet*  
[www.cbs.nl](http://www.cbs.nl)

Prijs: € 32,90 (exclusief verzendkosten)  
ISBN: 978-90-357-1526-4  
Oplage: 500

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen, 2010.  
Verveelvoudiging is toegestaan, mits het CBS als bron wordt vermeld.

# *Voorwoord*

Meer dan 70% van de sociale statistieken op het CBS worden gemaakt op basis van administratieve data. Meer en meer wordt ook in sociaalwetenschappelijk onderzoek gebruik gemaakt van dit soort data. Helaas heeft de ontwikkeling van een bijbehorende methodologie daarmee geen gelijke tred gehouden.

Voor u ligt een bundel met artikelen die de mogelijkheden en valkuilen beschrijven van sociaalwetenschappelijk onderzoek met behulp van administratieve data. Die mogelijkheden zijn legio, variërend van het vaststellen van daderprofielen in de criminologie tot de schatting van omzetten in macro-economisch onderzoek en van het vaststellen van de validiteit van enquêtegegevens tot het schatten van de loondaling ten gevolge van ontslag.

De valkuilen krijgen in deze bundel ook de aandacht die ze verdienen. Daarbij gaat het onder meer om de kwaliteit van administratieve gegevens, zoals problemen met de populatiedekking en problemen die veroorzaakt worden door de invoering van nieuwe en wijziging van bestaande registraties.

Deze problemen spelen zowel een rol bij de productie van statistieken als voor sociaalwetenschappelijk onderzoek. Daarom zijn we blij dat de Vrije Universiteit Amsterdam en het Centraal Bureau voor de Statistiek gezamenlijk deze bundel hebben samengesteld en daar ook artikelen aan hebben bijdragen.

Directeur-Generaal van de Statistiek  
Drs. G. van der Veen

Decaan Faculteit Sociale Wetenschappen  
Der Vrije Universiteit Amsterdam  
Prof. dr. A. Hemerijck

Den Haag/Heerlen/Amsterdam, december 2010



# Inhoud

<b>Voorwoord</b> .....	3
<b>Inleiding</b>	
<i>Bart F. M. Bakker en Léander Kuijvenhoven</i> .....	7
<b>Vaststellen van de validiteit van registervariabelen</b>	
<i>Bart F. M. Bakker</i> .....	15
<b>Participatie en loon na bedrijfseconomisch ontslag Kapitaalvernietiging of negatief signaal?</b>	
<i>Didier Fouarge, Ruben van Gaalen en Andries de Grip</i> .....	31
<b>Opleidingsniveau in registers: Een toets van de validiteit via loonfuncties</b>	
<i>Aslan Zorlu</i> .....	51
<b>Registers en criminologisch onderzoek</b>	
<i>Catrien Bijleveld</i> .....	65
<b>Nauwkeurigheid van gerapporteerde giften aan goededoelenorganisaties</b>	
<i>René Bekkers</i> .....	75
<b>Een buitenkansje! (Discussie over Bekkers, R.)</b>	
<i>Jelke Bethlehem</i> .....	89
<b>Koppeling van registers: Onmisbaar voor een beter inzicht in de volksgezondheid</b>	
<i>Anton E. Kunst</i> .....	93
<b>Continuïteit van zorg: Indicatoren voor lange termijn zorg en zekerheid in de GGz</b>	
<i>André Wierdsma</i> .....	111
<b>Het meten van de kwaliteit van administratieve bronnen: Recente resultaten en toekomstige ontwikkelingen</b>	
<i>Piet J. H. Daas, Sakia J. L. Ossen en Martijn Tennekkes</i> .....	127

<b>Het gebruik van BTW voor maand- en kwartaalomzetschattingen in Europa: Een vergelijking tussen Nederland en het Verenigd Koninkrijk</b>	
<i>Pieter Vlag, Henk van de Velden en Nino Mushkudiani</i> .....	143
<b>Modellen voor micro-integratie</b>	
<i>Jeroen Pannekoek</i> .....	155
<b>Discussie over Pannekoek, J.: Modellen voor micro-integratie</b>	
<i>Stef van Buuren</i> .....	169
<b>Registers: Onderzoeksagenda voor de toekomst</b>	
<i>Piet J. H. Daas, Léander Kuijvenhoven en Kees Zeelenberg</i> .....	173

# *Registers en sociaalwetenschappelijk onderzoek: een geslaagde combinatie?*

Bart F.M. Bakker (VU/CBS) en Léander Kuijvenhoven (CBS)

## *1. Inleiding*

Het gebruik van administratieve gegevens, ook wel registraties genoemd, voor sociaalwetenschappelijk onderzoek neemt toe. De ontwikkeling van de methodologie voor het gebruik van registraties houdt daarmee echter geen gelijke tred. Dat was voor de Vrije Universiteit en het Centraal Bureau voor de Statistiek de reden om samen een conferentie te organiseren. Op deze conferentie is door wetenschappers vanuit verschillende disciplines hun inhoudelijk vaak interessante onderzoek gepresenteerd. Tevens gingen zij in op de methodologische en praktische problemen die men tegenkwam tijdens het onderzoek. Deze bundel is daarvan het resultaat.

In dit hoofdstuk gaan we eerst kort in op de redenen van het toegenomen gebruik van registraties. Vervolgens bespreken we kort de verschillende bijdragen. We sluiten af met een voorlopig antwoord op de vraag of registraties en sociaalwetenschappelijk onderzoek wel een geslaagde combinatie zijn.

## *2. Voordelen van registraties*

De voordelen van het gebruik van registraties in plaats van enquêtes zijn legio. In de eerst plaats is de kwaliteit van de informatie uit enquêtes afgenomen. Dat wordt voornamelijk veroorzaakt door een toegenomen non-response die ook steeds selectiever is geworden (Stoop 2005; Cobben 2010). Vooral de selectiviteit is een probleem omdat die leidt tot vertekening in de schattingen.

In de tweede plaats is het houden van enquêtes duurder dan het verwerven van registraties. Registraties zijn goedkoper vanwege het feit dat ze toch al om andere redenen dan voor sociaalwetenschappelijk onderzoek worden bijgehouden. Het gebruik van deze informatie is overigens niet gratis. Voordat data uit registraties bruikbaar zijn voor onderzoek moeten ze vaak uitgebreid worden bewerkt. Die kosten dienen ook in de beschouwingen meegenomen te worden. Enquêtes zijn duurder geworden vanwege de invoering van de zogenaamde flexwet. Daardoor is enquêtering via relatief goedkope freelancers niet langer een reële optie.

In de derde plaats zijn door de ontwikkelingen in de informatietechnologie steeds meer registraties beschikbaar gekomen. In een aantal gevallen zijn bestaande pa-

pieren registraties gedigitaliseerd zoals de bevolkingsadministratie, in andere gevallen is een registratie volledig nieuw ontworpen zoals de Polisadministratie van het UWV en de Belastingdienst. Door de invoering van de Wet op het Centraal Bureau voor de Statistiek van 2004 zijn deze registraties voor het CBS kosteloos toegankelijk geworden. Het CBS stelt deze vervolgens onder bepaalde condities ter beschikking aan onderzoekers.

Een andere relevante ontwikkeling in de informatietechnologie is de invoering en steeds bredere toepassing van het Sofinummer, tegenwoordig BurgerServiceNummer geheten. Daardoor zijn de mogelijkheden vergroot om registraties onderling te koppelen waardoor meer bruikbare databestanden kunnen ontstaan.

De belangrijkste oorzaak van de toename is echter gelegen in de nieuwe onderzoeksmogelijkheden die registraties bieden (Borghans, 2007; Bakker, 2009). Er kan onderzoek worden gedaan naar heel kleine subpopulaties of zeldzame fenomenen, omdat registraties vaak de totale populatie van de registratie beschrijven. Een mooi voorbeeld is een hoofdstuk uit het recent verschenen proefschrift van Smits (2010). Daarin onderzocht zij de redenen waarom mensen tussen de 30 en 40 jaar bij hun ouders zijn gaan wonen, en vanuit welke specifieke behoeftes dat gebeurt. Dat heeft ze volledig met registerdata uitgevoerd. Het was longitudinaal onderzoek, waarbij gekeken werd of het kind naar de ouders verhuisde of andersom en welke invloed factoren als een echtscheiding van het kind, of van de ouder, inkomensveranderingen van ouder en het kind van invloed zijn.

Vaak is het overigens zo dat niet alle relevante vragen kunnen worden beantwoord met alleen registervariabelen. Dan kan het koppelen van een enquête nog wel eens helpen om aan de ontbrekende variabelen te komen. Als de informatie niet aanwezig is in een bestaande enquête kan een enquête specifiek voor het verzamelen van de aanvullende informatie worden ontworpen. Als de populatie een kleine categorie mensen betreft, kunnen de registraties als steekproefkader dienen waardoor de benadering heel efficiënt kan verlopen.

In enquêtes zijn onderwerpen als criminaliteit of het vóórkomen van specifieke ziektes lastig zonder vertekening te meten. Dat wordt voor een belangrijk deel veroorzaakt door sociaalwenselijk antwoordgedrag, maar ook kunnen geheugenefecten hiervoor zorgen. Deze informatie is goed in registraties opgenomen. Bij de vaststelling hiervan speelt sociaalwenselijk antwoordgedrag een veel kleinere rol. De mogelijkheden voor intergenerationeel onderzoek zijn toegenomen. De relaties tussen ouders en kinderen zijn voor een groot deel vastgelegd in de Gemeentelijke Basisadministratie (GBA). Zo kunnen registratieve data van ouders en kinderen worden verbonden. Een alternatief design is oudere enquêtegegevens te gebruiken voor de ouders en die te combineren met registratieve data van de kinderen van recenter datum.

Het is met registratieve data gemakkelijker om longitudinaal onderzoek te doen. Zonder mensen te benaderen kunnen ze worden gevolgd in hun levensloop. Bovendien heb je geen last van panelsterfte anders dan dat mensen daadwerkelijk



sterven of emigreren. Een mogelijkheid is ook om oudere enquêtegegevens te combineren met recentere registratieve data, zodat een lange periode kan worden beschreven.

Een prachtig voorbeeld van wat je kunt doen met koppelingen tussen ouders en kinderen en gevoelige longitudinale registerdata is het onderzoek van Besjes en Van Gaalen (2008). Zij onderzochten de invloed van criminaliteit van vaders en moeders op de criminaliteit van hun kinderen, ook weer volledig met registerdata. Er kwam overigens een sterk verband uit, zeker voor de moeders en hun kinderen. Relatief eenvoudig kunnen ad hoc vragen worden beantwoord als de informatie in registraties aanwezig is, ook als die vragen betrekking hebben op het verband tussen variabelen die men niet snel samen in een enquête tegenkomt. Dat is vooral voor het CBS belangrijk om die vragen uit de samenleving te kunnen beantwoorden, want daar is het CBS onder andere voor. Het voorbeeld dat het meest wordt gebruikt is dat van het verband tussen voortijdig schoolverlaten en criminaliteit.

### 3. *In deze bundel*

Een van de aspecten van de kwaliteit van data is de validiteit ervan. Als de validiteit van een meting laag is, dan wordt niet gemeten wat de bedoeling was en wordt ook wel gesproken van vertekening. Het is echter niet zo eenvoudig om de validiteit van metingen daadwerkelijk vast te stellen. In de bijdrage van Bakker wordt een methode gepresenteerd waarmee de constructvaliditeit kan worden vastgesteld. Hij gebruikt daarvoor Lineaire structurele modellen waarbij iedere variabele twee keer wordt gemeten: eenmaal in een enquête en eenmaal in een register. Dit model past hij toe op leeftijd, geslacht, opleidingsniveau en uurloon uit het OSA-aanbodpanel (enquête) en het Sociaal Statistisch Bestand (SSB, register). Hij toont aan dat de validiteit van metingen van leeftijd en geslacht in het register en in de enquête goed zijn. Die van opleidingsniveau en uurloon zijn ook behoorlijk, maar wel duidelijk minder. Bovendien is de validiteit van het opleidingsniveau uit het register iets hoger dan die uit de enquête.

Fouarge, Van Gaalen en De Grip gaan in hun bijdrage in op de vraag of ontslag leidt tot een vermindering van de waarde op de arbeidsmarkt. Zij gebruiken daarbij administratieve ontslaggegevens en combineren dat met loopbaangegevens uit het SSB. Daarmee is dat een goed voorbeeld van longitudinaal onderzoek dat relatief goedkoop door de combinatie van een groot aantal registraties kan worden uitgevoerd. Zij komen tot de conclusie dat de mensen die worden ontslagen om bedrijfseconomische redenen inderdaad een minder goede positie op de arbeidsmarkt innemen. Weliswaar is er na verloop van enkele jaren vrijwel geen effect meer op de kans op werk, maar het loon is nog wel lager.

Zorlu constateert dat het ontbreken van een variabele opleidingsniveau in administratieve bronnen gevolgen heeft voor de bruikbaarheid van registraties. Hij ver-

welkomt dan ook het initiatief van het CBS om de wel aanwezige informatie uit relatief recente administratieve bronnen te combineren met vele jaargangen van de Enquête beroepsbevolking (EBB). Hij onderzoekt de validiteit van deze informatie voor gebruik in onderzoek naar loonverhoudingen onder allochtonen. In het bijzonder de eerste generatie allochtonen ontbreken dikwijls in de registers. Daarnaast zijn de eerste en tweede generatie allochtonen ondervertegenwoordigd in de EBB. Toch concludeert hij dat de informatie redelijk bruikbaar is, maar dat zorgvuldig moet worden gekeken naar de interpretatie van de schattingen in loonvergelijkingen.

Bijleveld stelt dat registers een grote potentie en waarde hebben voor het verrichten van criminologisch onderzoek. In haar bijdrage richt zij zich vooral op het maken van daderprofielen. Zij geeft een overzicht van verschillende typen registers, zowel registers die specifiek gebruikt worden voor criminologisch onderzoek, maar ook registers van meer algemene aard. Belangrijk is het op te merken dat zij vaak werkt met nog niet bewerkte registers. Op deze manier staat zij dichtbij de bron. Aan de hand van een aantal casussen onderzoekt zij, soms op anekdotische wijze, waar de zwakke en sterke punten van registers liggen.

Het artikel van Bekkers is een voorbeeld van een studie waarin registratieve gegevens als standaard worden gebruikt om de kwaliteit van enquêtegegevens aan af te meten. In de bijdrage van Bekkers wordt de door respondenten in een enquête gerapporteerde giften aan Greenpeace vergeleken met de giften die bij de organisatie geregistreerd zijn als afkomstig van de huisadressen van deze respondenten. Een deel van de respondenten verzuimt giften die zijn gedaan te rapporteren in de enquête. Een bijna even groot deel van de respondenten rapporteert giften gedaan te hebben die niet zijn geregistreerd. De gemiddelde vertekening is daardoor nagenoeg nul. De verschillen tussen zelfrapportage en registratie blijkt groter te zijn onder ouderen en tweeverdieners. Dit kan verklaard worden door geheugenproblemen en gebrek aan informatie. De verschillen zijn gering onder respondenten die sociaal wenselijk gedrag rapporteren zoals stemgedrag of bloeddonatie. Bethlehem reageert daar kritisch op door te stellen dat Bekkers ten onrechte aanneemt dat de registratie foutloos is.

De bijdrage van Kunst is een overzicht van de mogelijkheden die koppeling van registers bieden voor onderzoek op het terrein van de volksgezondheid. De bijdrage begint met een overzicht van ontwikkelingen in onderzoek naar sterfte in Nederland. Dit overzicht illustreert dat het onderzoek naar volksgezondheid in Nederland zich in de afgelopen jaren in snel tempo heeft verbreed en verdiept. Koppeling van registers biedt vooral nieuwe mogelijkheden voor differentiatie naar persoonskenmerken, voor decompositie van gefaseerde processen en voor longitudinaal onderzoek naar causale relaties. Verder stelt hij dat doordat registers vaak zijn samengesteld uit grote aantallen observaties een veel grotere precisie en detaillering kan worden bereikt dan in persoonsenquêtes of traditioneel epidemiologisch onderzoek. Hij blijft echter wel kritisch en laat met een concreet voorbeeld

waarbij de Landelijke Medische Registratie (LMR) gebruikt wordt, zien dat in sommige gevallen praktische problemen de gebruikswaarde van registers kunnen verminderen. Voor veel onderzoek blijven persoonsenquêtes een belangrijke bron van informatie, dat overigens wel verrijkt kan worden door koppeling aan registers.

Wierdsma gaat in op de mogelijkheden om met administratieve bronnen de mate van continuïteit in gebruik van de geestelijke gezondheidszorg vast te stellen. Dat valt niet mee, omdat de informatie longitudinaal moet zijn, alle belangrijke zorgfuncties moet omvatten, relevante uitkomstmaten in beeld moet brengen, standaard meetinstrumenten moet gebruiken, ook relevante verklarende variabelen moet vastleggen, en diverse settings of zorgorganisaties moet omvatten. Psychiatrische casusregisters zijn in staat om verschillende typen van zorgcontinuïteit longitudinaal en integraal te monitoren. Het zicht op andere dan GGz-instellingen is echter minder. Verder gaat hij in op de bruikbaarheid van een aantal continuïteitsmaten. De inventarisatie van indicatoren van zorgpatronen laat zien dat de verschillende typen continuïteit kunnen worden uitgedrukt in standaardmaten.

Daas et al. presenteren een op het Centraal Bureau voor de Statistiek (CBS) ontwikkeld kwaliteitsraamwerk dat het mogelijk maakt de statistische bruikbaarheid van administratieve en andere secundaire bronnen in kaart te brengen. Dit kwaliteitsraamwerk bestaat uit drie hoogdimensionale manieren van kijken (niveaus) naar de kwaliteit van een databron. In de eerste twee niveaus, die Bron en Metadata zijn genoemd, wordt gekeken naar kwaliteitsaspecten die met de levering van de bron en de beschikbaarheid en duidelijkheid van de meta-informatie van de bron samenhangen. Problemen op één of beide niveaus kunnen ervoor zorgen dat het CBS niet of slechts beperkt van de gegevens in een bron gebruik kan maken. Voor het bepalen van de kwaliteitsaspecten in Bron en Metadata is een checklist ontwikkeld. De resultaten van de checklist worden besproken voor een aantal bronnen. Voor het bepalen van de kwaliteit van de aspecten die tot het derde gezichtspunt behoren, het Data-niveau, is een checklist niet echt toepasbaar. Hiervoor dient een andere methode te worden gebruikt. Een aantal mogelijkheden wordt besproken, waarbij de aandacht vooral gericht is op de verschillende groepen van indicatoren en visualisatie-technieken.

Vlag et al. gaan in op het gebruik van BTW-gegevens voor het schatten van de omzetten in bedrijfseconomische statistieken. Belangrijk daarbij is dat in geheel Europa gezocht wordt naar mogelijkheden om voor deze statistieken meer gebruik te maken van administratieve gegevens en minder van enquêtering. Zij vergelijken de mogelijkheden voor Nederland en het Verenigd Koninkrijk. Hoewel de aangiftesystematiek in deze landen verschilt, blijkt het toch mogelijk om grotendeels dezelfde methode te volgen om tot uitkomsten te komen. Naar verwachting kan deze methode ook in andere landen worden toegepast.

Pannekoek geeft een voorbeeld van het combineren en integreren van administratieve data met enquêtegegevens op microniveau. Hij bespreekt een aantal verschil-

lende modellen die behulpzaam zijn bij het integreren van de soms conflicterende informatie uit de verschillende databronnen en de bekende relaties waaraan de variabelen bij economische statistieken moeten voldoen. Daarbij gaat het om het oplossen van het probleem hoe te komen tot schattingen voor het binnenwerk van een tabel wanneer de marginalen bekend zijn. Van Buuren noemt de bijdrage van Pannekoek een welkome aanvulling op deze methodes, maar heeft op enkele technische aspecten wat aarzeling.

Daas et. al. schetsen een onderzoeksprogramma op het terrein van registers voor de toekomst. Het onderzoeksprogramma richt zich vooral op het statistische proces van het CBS. Ook besteden ze meer aandacht aan methodologische vraagstukken en wat minder op vraagstukken dat gemotiveerd wordt vanuit de inhoud. Dat betekent echter niet dat dit geen interessante vraagstukken oplevert voor de sociale wetenschappen. Hun bespreking van onderzoeksvragen van het statistische proces laat zien dat er juist bijzonder interessante onderzoeksonderwerpen zijn voor hen. Op sommige gebieden wordt er nu al gezamenlijk onderzoek verricht en ze roepen op tot een nauwere samenwerking.

#### 4. *Registers en sociaalwetenschappelijk onderzoek: een geslaagde combinatie?*

De bijdragen aan deze bundel laten duidelijk zien dat er veel mogelijkheden zijn voor het gebruik van registraties in sociaalwetenschappelijk onderzoek. Tegelijkertijd is duidelijk dat het werken met deze data niet eenvoudig is. Samengevat zijn er de volgende problemen. Voor een gedetailleerder overzicht verwijzen we naar Bakker (2009).

Het vereist inzicht in de belangen van en de administratieve praktijken van registratiehouders om een strategie te ontwerpen op welke wijze administratieve gegevens gebruikt kunnen worden. Helaas hebben registerhouders vaak andere belangen dan het vervaardigen van statistieken en doen van onderzoek. Ook is het goed om na te denken over de belangen die (groepen) geregistreerden hebben bij het op een bepaalde manier geregistreerd staan. Zo bestaat in vrijwel alle landen waarin een Volkstelling wordt gehouden op basis van registraties het probleem dat een groot deel van de studenten niet op hun geregistreerde adres wonen. Daar hebben ze vaak financiële belangen bij.

Hetzelfde soort inzicht is ook nodig om registratieve data longitudinaal te gebruiken. Vaak ontbreken stukjes in administraties die onder bepaalde aannames kunnen worden opgevuld. Zo ontbreken in de Polisadministratie van het verslagjaar 2006 ongeveer 2,8 miljoen baanperiodes van een maand of vier weken. Als deze informatie niet zou worden aangevuld zouden er volgens de gegevens onwaarschijnlijke aantallen personen zijn die hun baan kwijtraken en later weer terug krijgen.

Verder is het noodzakelijk om goed onderzoek te doen naar de dekking van de

populatie. Vaak ontbreken delen van de populatie in de registratie en meestal is dat selectief. Voor de uitkomsten van onderzoek kan dat flinke vertekeningen opleveren. Ook is het van belang om steeds te beoordelen of het koppelingsrendement van bestanden voldoende is om het onderzoek mee uit te voeren. Daarbij is de mate van selectiviteit belangrijker dan het rendement zelf, naar analogie met non-respons in enquêtes. Mocht de selectiviteit groot zijn, dan is het meestal verstandig om de niet-gekoppelde records in het onderzoek te blijven betrekken.

Een praktisch probleem kan ook de omvang van de bestanden zijn, omdat het meestal gaat om bestanden met miljoenen records. Verreweg de meeste hardware en software is er op gebouwd om relatief kleine bestanden te verwerken. Het is goed om na te denken over vormen van datareductie die wel de informatiewaarde volledig in tact laat, maar die ervoor zorgt dat programmatuur snel tot resultaten leidt. Een eerste idee daarvoor is de data te reduceren door een kruistabel van de te onderzoeken variabelen te maken en deze op te slaan in een formaat waarbij de celvulling als wegingsfactor wordt opgenomen. Een tweede idee is de data reduceren door eerst een correlatie- of covariantiematrix te berekenen en deze als input te gebruiken voor verdere analyses. Verder nemen ook de mogelijkheden van de informatietechnologie toe en wordt deze technologie steeds goedkoper, waardoor het breder ingezet kan worden. Dit maakt het mogelijk dat meer computers worden ingezet om grote bestanden te analyseren. Een voorbeeld van deze technologie is een computercluster. Een computercluster bestaat uit een netwerk van computers die als het ware een berekening als één enkele computer lijken uit te voeren. Buiten het feit dat grote bestanden gebruikt kunnen worden zal ook de berekeningen veel sneller verlopen dan als men hiervoor een enkele computer zou gebruiken.

Uit bovengenoemde problemen blijkt wel dat er nog heel wat methoden en technieken ontwikkeld moeten worden. Toch kan het huwelijk tussen registerdata en sociaalwetenschappelijk onderzoek alvast aangekondigd worden. De mogelijkheden die deze data bieden zijn te talrijk en te veelbelovend om daar geen gebruik van te maken.

## *Referenties*

Bakker, Bart F.M. (2009). *Trek alle registers open!* (Amsterdam: Vrije Universiteit)

Besjes, G. en R. van Gaalen, Jong geleerd, fout gedaan? Een onderzoek naar de intergenerationele overdracht van criminaliteit met gegevens uit het Sociaal Statistisch Bestand, In *Bevolkingstrends*, jrg. 56, nr. 2, blz. 23–31

Borghans, L. (2007). Anders denken, In: *Edata&research*, jrg. 2, nr 2, blz. 8

Cobben, F. (2010). *Nonresponse in Sample Surveys, Methods for Analysis and Adjustment*. *Proefschrift Universiteit van Amsterdam* (Den Haag / Heerlen: CBS)

Smits, A. (2010). *The role of parents in their adult children's housing and residential locations*. Proefschrift Universiteit van Amsterdam (Amsterdam: Universiteit van Amsterdam)

Stoop, I. (2005). *The hunt for the last respondent. Nonresponse in sample surveys* (Den Haag: SCP)

# *Vaststellen van de validiteit van registervariabelen*

B.F.M. Bakker (VU/CBS)<sup>1)</sup>

## *1. Inleiding*

Het gebruik van administratieve gegevens, ook wel registraties genoemd, voor sociaalwetenschappelijk onderzoek neemt toe. De redenen die daarvoor zijn aan te voeren zijn legio. De kwaliteit van de informatie die verkregen wordt uit enquêtes is minder geworden, voornamelijk vanwege de toename van en de toegenomen selectiviteit van non-response. Enquêteren is bovendien duurder dan het gebruiken van al aanwezige registraties. Door ontwikkelingen in de informatietechnologie zijn steeds meer registraties beschikbaar gekomen. Bovendien is door de invoering van het Sofinummer, tegenwoordig Burgerservicenummer geheten, de mogelijkheden vergroot om registraties onderling te koppelen waardoor meer bruikbare databestanden ontstaan.

Ondanks de toename van het gebruik van registratieve gegevens in sociaalwetenschappelijk onderzoek, is nog maar weinig bekend over mogelijke kwaliteitsproblemen ervan. Lang is gedacht dat het samenstellen van statistieken of het beantwoorden van sociaalwetenschappelijke vragen met behulp van registraties een kwestie was van tellen. Steeds vaker wordt erkend dat dat een naïeve gedachte is (zie bijvoorbeeld Al & Bakker, 2000; Statistics Denmark, 1994; Statistics Finland, 2005; Wallgren & Wallgren, 2007; Bakker, 2009a). Een belangrijke vraag is of de te meten concepten ook valide gemeten kunnen worden met behulp van registraties: meet je dat wat beoogd wordt te meten of zijn daar systematische afwijkingen in terug te vinden.

Een eerste poging om in kaart te brengen welke fouten onderscheiden zouden moeten worden in registraties is opgenomen in Bakker (2009a). Analooq aan Groves et al. (2004) wordt een onderscheid gemaakt tussen meet- en representatiefouten. Bij meetfouten gaat het om metingen van het begrip of concept en bij de andere fouten gaat het om bij wie die begrippen of concepten gemeten worden. Hoewel er een zekere samenhang kan bestaan tussen beide soorten fouten, zijn de representatiefouten niet van belang voor het vaststellen van de validiteit van registervariabelen.

Op de meting van variabelen in registraties heeft de onderzoeker in de praktijk vrijwel geen invloed. Dat betekent dat het administratieve begrip het uitgangspunt is. Soms kan door gebruik te maken van verschillende variabelen die in gecombi-

neerde registraties aanwezig zijn, het sociaalwetenschappelijke concept worden benaderd. Ontbreken die gegevens, dan moet noodgedwongen met het administratieve concept worden gewerkt. Voor de validiteit kan dat negatieve gevolgen hebben.

Administratieve vertraging is een andere belangrijke foutenbron in registers die bij kan dragen aan een lagere validiteit. Daarvan is sprake als een gebeurtenis later wordt geregistreerd dan hij heeft plaatsgevonden. Het duidelijkste voorbeeld is dat van de vertraging in de registratie van huwelijken van allochtonen die trouwen in hun land van herkomst. Vaak wordt pas als de huwelijkspartner overkomt het huwelijk in de GBA geregistreerd. Soms gaat daar een periode van 1 à 2 jaar overheen. Ook dit kan voor systematische vertekening zorgen.

In de bovengenoemde literatuur wordt het probleem met de validiteit van metingen ruimschoots erkend. Tegelijkertijd wordt de mate waarin afgeweken wordt van het werkelijke concept alleen in kwalitatieve zin gegeven. In dit artikel wordt een methode gepresenteerd om de validiteit van registervariabelen te kwantificeren. Daarbij wordt gebruik gemaakt van de klassieke testtheorie die ontwikkeld is binnen de psychologie om de validiteit en betrouwbaarheid van psychologische tests te bepalen.

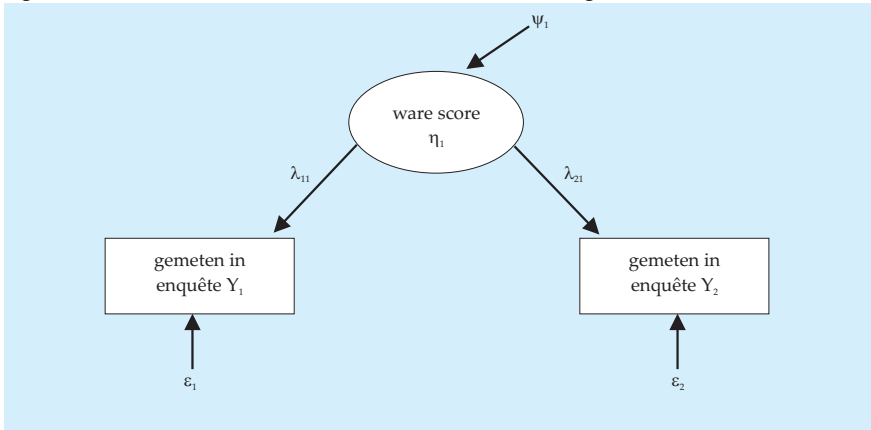
## 2. *Validiteit en betrouwbaarheid*

De kwaliteit van de meting van een variabele kan met verschillende indicatoren worden bepaald. De belangrijkste daarvan zijn betrouwbaarheid en validiteit. Bij betrouwbaarheid gaat het om toevallige fouten. Als men verschillende keren waarneemt bij dezelfde eenheden, is dat een manier om deze toevalligheden op het spoor te komen. Hetzelfde kan door meerdere waarnemers dezelfde waarneming te laten verrichten. (zie bijvoorbeeld 't Hart et al., 2005; McCall, 2001), De betrouwbaarheid van een meetinstrument kunnen we vaststellen door deze (ten minste) twee maal te gebruiken bij dezelfde (steekproef)populatie. In figuur 1 is het model weergegeven dat dan wordt geschat. In het ovaal is de ware score  $\eta_1$  opgenomen. Dit wordt ook wel de latente variabele genoemd, waarmee het gewenste concept wordt weergegeven. Deze wordt gemeten door twee maal het meetinstrument af te nemen. Het product van de  $\lambda_{11}$  en de  $\lambda_{21}$  is de waargenomen correlatie tussen de metingen op de beide tijdstippen en wordt ook wel de test-hertest-correlatie genoemd. De  $\epsilon_1$  en de  $\epsilon_2$  zijn de meetfouten. Dit model kan trouwens niet worden uitgerekend, omdat er meer onbekenden zijn dan vergelijkingen. Door aan te nemen dat  $\lambda_{11}$  en  $\lambda_{21}$  aan elkaar gelijk zijn, kan dat wel.

Bij validiteit gaat het om systematische fouten. Als een meetinstrument van een variabele niet nauwkeurig de conceptuele variabele meet maar eigenlijk iets anders, dan is de variabele niet valide gemeten. We spreken dan ook wel van vertekening. Er zijn verschillende manieren om de validiteit vast te stellen en te kwantifi-



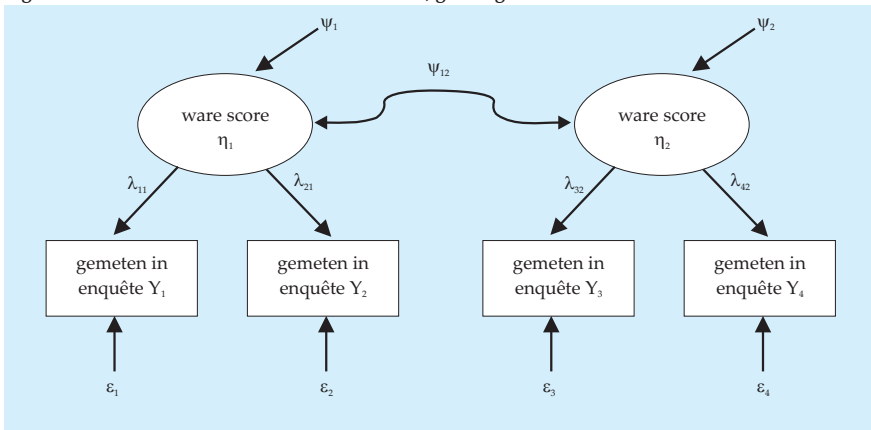
**Figuur 1. Vaststellen van de betrouwbaarheid door herhaalde meting**



ceren (McQueen en Knussen, 2002:95–98; 't Hart et al., 2005:187–188).

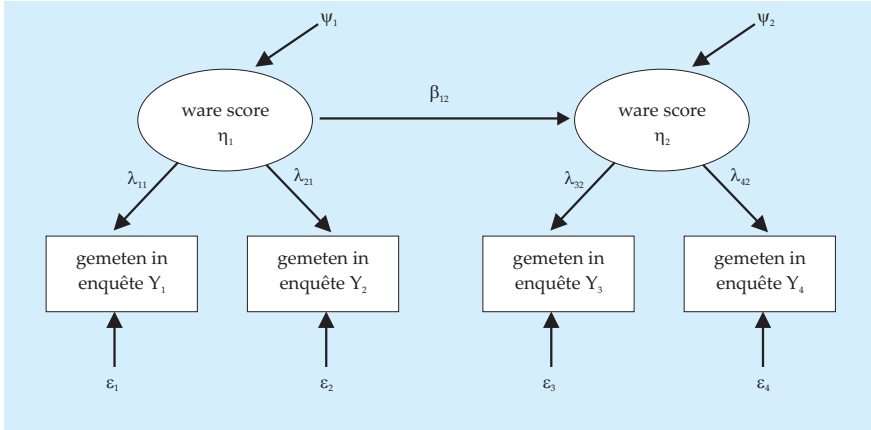
In de eerste plaats gaat het dan om de zogenaamde kruisvaliditeit of soortgenootvaliditeit. Daaronder wordt de samenhang met een verwante variabele verstaan. Als we bijvoorbeeld weten dat een variabele goed gemeten is met een bepaald meetinstrument in een enquête, dan kan de kruisvaliditeit van een variabele met een ander meetinstrument, gecorrigeerd voor de mate van betrouwbaarheid van beide meetinstrumenten, met behulp van het model in figuur 2 worden geschat. Daarbij wordt bij dezelfde (steekproef)populatie beide meetinstrumenten twee keer afgenomen. De  $\psi_{12}$  is dan de kruisvaliditeit: de voor onbetrouwbaarheid gecorrigeerde correlatie tussen de ware scores  $\eta_1$  en  $\eta_2$ . De producten van  $\lambda_{11}$  en  $\lambda_{21}$  en van  $\lambda_{32}$  en  $\lambda_{42}$  zijn weer de test-hertest-correlaties van respectievelijk de meetinstrumenten 1 en 2. De  $\epsilon_1$  en  $\epsilon_2$  en de  $\epsilon_3$  en  $\epsilon_4$  zijn de metingen voor onbetrouwbaarheid.

**Figuur 2. Kruisvaliditeit van een meetinstrument, gecorrigeerd voor meetonbetrouwbaarheid**



Een tweede soort validiteit is de predictieve validiteit. Daarmee wordt bedoeld dat het meetinstrument de variabele zo moet meten dat hij een andere variabele goed voorspelt. Samen met de kruisvaliditeit wordt dit wel criteriumvaliditeit genoemd. Als we weer corrigeren voor de onbetrouwbaarheden van de beide meetinstrumenten, ziet het model eruit als in figuur 3.

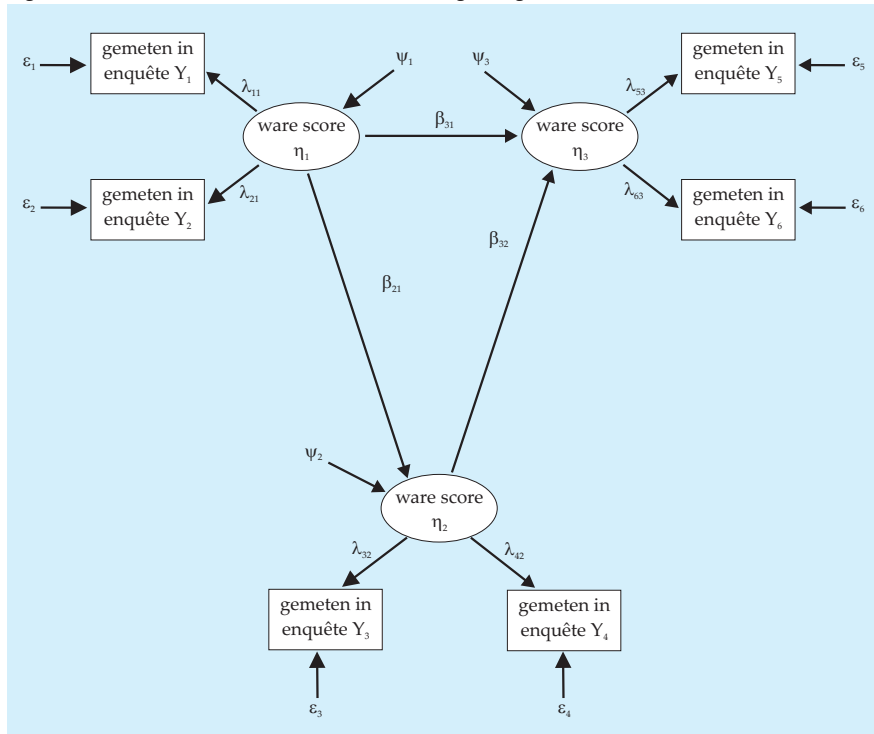
**Figuur 3. Predictieve validiteit van een meetinstrument, gecorrigeerd voor meetonbetrouwbaarheid**



De laatste vorm van validiteit die ik bespreek is die van de constructvaliditeit. Volgens Singleton en Straits (2005) gaat het er bij de meting van deze vorm van validiteit om te formuleren met welke variabelen men theoretisch een samenhang verwacht en met welke niet. Vervolgens test men deze theoretische verwachtingen empirisch. Hoe vaker de hypothesen niet worden verworpen, hoe hoger de validiteit van het construct. 't Hart et al. (2005) stellen dat deze vorm van validiteit in zekere zin de eerder genoemde vormen van validiteit als kruis- en predictieve validiteit omvat, maar dat daarenboven constructvaliditeit geworteld is in een theorie. Binnen de klassieke testtheorie wordt de validiteit van meerdere constructen vaak simultaan vastgesteld. Daarbij gaat het erom dat meer dan twee constructen samenhangen volgens verwachting. Als het beperkt is tot twee constructen is het ook wel bekend onder de naam convergentievaliditeit. Als we weer corrigeren voor de meetonbetrouwbaarheid van de variabelen, ziet het model er uit zoals weergegeven in figuur 4.

Uiteraard is een noodzakelijke voorwaarde voor het vaststellen van de validiteit van een meetinstrument dat we *weten* op welke wijze de constructen met elkaar samenhangen. Dat geldt voor de kruisvaliditeit, de predictieve en de constructvaliditeit. Dit referentiekader is lang niet altijd aanwezig. Vaak is er niet meer dan een theoretisch idee op welke wijze variabelen samenhangen. Als dat het geval is, kunnen meerdere meetinstrumenten worden vergeleken. Die meetinstrumenten die

Figuur 4. Constructvaliditeit van meetinstrumenten gecorrigeerd voor de meetonbetrouwbaarheid



het meest in overeenstemming zijn met de theoretisch verwachte samenhangen hebben dan de hoogste validiteit.

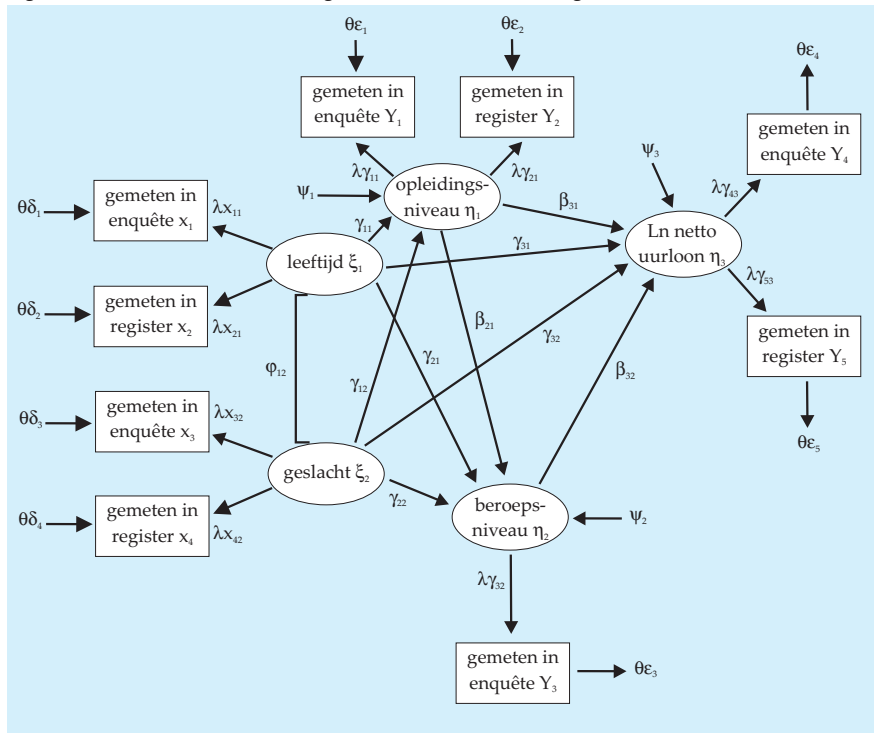
Tot nu toe gingen we ervan uit dat we alle variabelen twee maal via dezelfde vraagstelling in een enquête hebben gemeten. Daarmee kan worden vastgesteld of de vraagstelling betrouwbare resultaten oplevert. Het ligt dan voor de hand om aan te nemen dat de  $\lambda$ 's die bij dezelfde vraagstelling behoren aan elkaar gelijk te stellen. Dat hoeft overigens niet. Een vraagstelling kan bij herhaling toch een ander resultaat opleveren. Dit kan bijvoorbeeld veroorzaakt worden door een verschillende context of door geheugeneffecten. Dit is bijvoorbeeld het geval als de vragen binnen hetzelfde vraaggesprek worden herhaald, de respondent gedurende het vraaggesprek nog nagedacht heeft over het antwoord en bij de herhaling tot een ander antwoord komt.

Het is ook mogelijk dat er verschillende vraagstellingen worden gebruikt waarvan verondersteld wordt dat beide het gewenste concept meten. Het ligt dan veel minder voor de hand dat de twee verschillende vraagstellingen het concept even goed meten. Dit komt tot uitdrukking in een verschil in de  $\lambda$ 's die bij hetzelfde latente variabele behoren en uiteraard ook bij een verschil in de  $\varepsilon$ 's. Die krijgen dan echter een andere betekenis dan een meting van de meetonbetrouwbaarheid: het wordt dan een meting van de mate waarin het latente concept wordt gemeten en kan als

zodanig dus als een meting van de validiteit worden beschouwd. Idealiter zou je ieder van de vraagstellingen twee keer moeten afnemen om weer de betrouwbaarheid en validiteit te kunnen onderscheiden.

Hoe kunnen we nu gebruik maken van deze modellen om de validiteit van registervariabelen vast te stellen? In plaats van twee verschillende vraagstellingen in een enquête op te nemen, kun je ook een enquête koppelen aan een register en zo een meting uit een enquête vergelijken met de meting uit een register. Daarvoor is bijvoorbeeld het loonverwervingsmodel bruikbaar zoals weergegeven in figuur 5. In het model zijn vijf latente variabelen opgenomen: leeftijd, geslacht, opleidingsniveau, beroepsniveau en de logaritme van het netto-uurloon. Alle variabelen met uitzondering van het beroepsniveau zijn gemeten in een enquête en in een register. Aangetekend moet worden dat de variabele opleidingsniveau niet in zijn geheel een registervariabele is. Het betreft een combinatie van register- en enquête-informatie. Het grootste deel van de mensen heeft

**Figuur 5. Het model voor de schatting van de validiteit van vier registervariabelen**



echter een opleidingsniveau vanuit een administratief bestand toegekend gekregen. We verwachten dat de effecten van de meest valide meting van opleidingsniveau op uurloon het grootst is, gecontroleerd voor leeftijd, geslacht en beroepsniveau. Meetfouten leiden immers in de meeste gevallen tot lagere correlaties en daarmee tot lagere effecten. Aangezien hier het effect niet voor de afzonderlijke variabelen

wordt vastgesteld, maar van de gemeenschappelijke variantie van de twee metingen, vertaalt dit zich in een lagere meetfout voor de meest valide metingen. Het model wordt uitgerekend voor mensen die in loondienst werkzaam zijn voor 12 uur of meer per week. Daarmee worden kleine baantjes waarvoor de samenhangen wel eens anders kunnen liggen, uitgesloten.

### 3. *De data*

#### 3.1 *Enquêtedata*

We maken voor wat betreft de enquêtedata gebruik van het zogenaamde OSA-Aanbodpanel 2004. In het databestand zijn 4 785 respondenten opgenomen. De steekproef is een huishoudensteekproef gestratificeerd naar leeftijd, geslacht, regio en huishoudensamenstelling. Personen die jonger zijn dan 16 of ouder dan 64 jaar of personen die volledig dagonderwijs volgen zijn uitgesloten. De eerder deelnemende huishoudens zijn opnieuw benaderd. Als een persoon een huishouden verlaat wordt het gehele nieuwe huishouden benaderd. Personen die buiten de doelpopulatie vallen worden niet langer gevolgd. Personen die bij een eerdere golf niet maar inmiddels wel tot de doelpopulatie vallen, worden wel ondervraagd. De data zijn vervolgens zodanig herwogen dat de steekproef de verdeling van bovengenoemde variabelen representeert zoals die in de Enquête beroepsbevolking worden gemeten (zie verder NIWI, 2010). De vraaggesprekken zijn gehouden rond de datum van 1 oktober 2004. Dat wordt beschouwd als de peildatum van deze OSA-golf.

*Leeftijd* is in de enquête gemeten door de geboortedatum te vragen en deze vervolgens om te rekenen naar een leeftijd op de enquêtedatum. *Geslacht* is gemeten met de vraag: "Wat is uw geslacht?"

*Opleidingsniveau* is gemeten met de vraag: "Wat is de hoogste opleiding die u heeft voltooid, waar u een diploma voor heeft behaald?" De respondent krijgt dan een kaart voorgelegd waarop 40 opleidingen genoemd worden. Deze opleidingen stammen uit een lange periode om alle generaties voldoende keuzemogelijkheden te bieden. Enkele voorbeelden daarvan zijn: "lagere school, basisschool", "LTS: Lagere Technische School", "MBO: Middelbaar Beroepsonderwijs", "Gymnasium", "Primair leerlingwezen", "BOL: assistent-opleiding", "Dagopleiding of deeltijdopleiding in het buitenland". Als het beroepsonderwijs of een buitenlandse opleiding betreft, wordt gevraagd: "Welk type school en studierichting volgde u in het laatste jaar daarvan?". Deze informatie is later omgezet in een code van de Standaard Onderwijs Indeling.

*Beroepsniveau* is vastgesteld met een uitgebreide vraagstelling waarin onder andere naar de belangrijkste werkzaamheden, het aantal personen waarover leiding wordt gegeven, en de belangrijkste leidinggevende werkzaamheden is gevraagd. De informatie is vervolgens gecodeerd in de Standaard Beroepenclassificatie 1992 (CBS, 1993). Het *ln netto-uurloon* is vastgesteld door de vraag: "Kunt u mij zeggen hoeveel uw

netto maandloon bedraagt?”, waarbij de interviewer eerst noteert of de respondent een bedrag per week, per vier weken, per maand of per jaar noemt en vervolgens het desbetreffende bedrag. Verder is de vraag gesteld naar hoeveel uren men volgens het contract werkt. Deze informatie heeft steeds betrekking op de hoofd baan. Door de informatie te combineren is het uurloon berekend, waarna dit logaritmisch is getransformeerd.

### 3.2 *Registerdata*

De registerdata zijn ontleend aan het Sociaal Statistisch Bestand (SSB). Dit is een stelsel van koppelbare registers en enkele (steekproef)enquêtes, waarvan de definitieve versie zodanig is opgeschoond dat tot op grote hoogte consistentie in de data is bereikt (Arts & Hoogteijling, 2002; Bakker, 2008). De methodologie die daarbij gebruikt wordt is micro-integratie (Bakker, 2009a). Door micro-integratie wordt de kwaliteit van de data verhoogd, omdat deze worden geharmoniseerd, gecompliceerd en ontdaan van meetfouten. Daarbij wordt een stelsel van beslisregels toegepast. Daarna worden de bestanden register genoemd. In het onderstaande bespreken we de variabelen en geven daarbij ook in grote lijnen aan welke beslisregels daarop van toepassing zijn geweest.

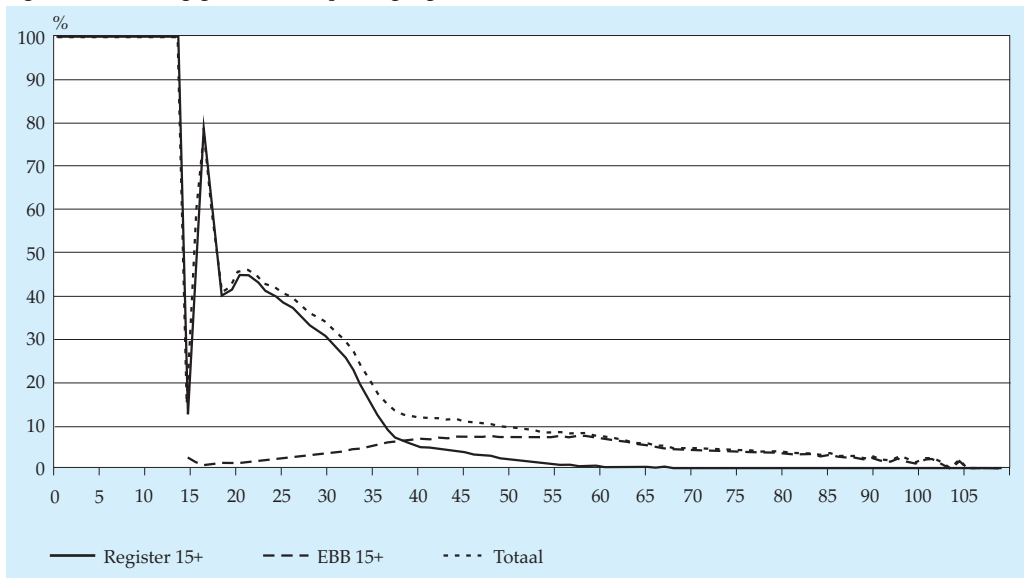
*Leeftijd* en *geslacht* worden ontleend aan de Gemeentelijke BasisAdministratie (GBA). Deze wordt voor demografische variabelen kwalitatief beter geacht dan dezelfde variabelen uit andere bronnen. Mochten personen niet voorkomen in de GBA dan worden hun leeftijd en geslacht ontleend aan andere bronnen. In het databestand dat voor het schatten van het model wordt gebruikt, worden alleen gekoppelde records gebruikt die daarom per definitie ook in de GBA voorkomen.

Voor *opleidingsniveau* is geen registratie beschikbaar waarmee de Nederlandse bevolking wordt gedekt. Wel is informatie opgenomen in verschillende registraties van scholieren en studenten zoals de inschrijvingen in het hoger onderwijs (sinds 1983), het Examen Resultaten Register (sinds 1999), de onderwijsnummerbestanden voortgezet onderwijs (vanaf 2004), de onderwijsnummerbestanden middelbaar beroeps-onderwijs (vanaf 2005) en enkele kleinere registraties. Naast deze administratieve bronnen worden de gegevens uit de Enquête beroepsbevolking (EBB) gebruikt van de jaargangen 1996–2008. Daarvan wordt getoetst of het opleidingsniveau op het peilmoment nog geldig is. Als de kans groter is dan 5% dat het opleidingsniveau niet geldig is, wordt het record van de EBB verwijderd. Bij opleidingsniveau wordt altijd de hoogste opleiding genomen die in de bronnen is terug te vinden. Als twee bronnen een gelijk opleidingsniveau laten zien, wordt de opleidingsrichting gekozen uit de registraties. Dat laatste is niet van belang voor dit onderzoek, omdat alleen het opleidingsniveau wordt gebruikt. Verder wordt aangenomen dat kinderen van 12 jaar en jonger nog geen diploma in het voortgezet onderwijs hebben behaald. Voor zover 12–14-jarigen niet in een onderwijsnummerbestand voorkomen, wordt hetzelfde verondersteld. Zo wordt in totaal ongeveer van 7 miljoen mensen het opleidingsniveau vastgesteld. Van deze informatie wordt op verschillende peilmomen-

ten het opleidingsniveau bepaald. Hier wordt gebruik gemaakt van de informatie over het opleidingsniveau op ultimo september 2004. Dat is nagenoeg gelijk aan het peilmoment van het OSA-Aanbodpanel 2004.

Kenmerk van de registraties met onderwijsinformatie is dat ze informatie van recente datum bevatten. Dat betekent dat de registerinformatie selectief is naar leeftijd en alle kenmerken die samenhangen daarmee. Een andere beperking is dat in de onderwijsregistraties alleen het bekostigd onderwijs is opgenomen. Bedrijfsopleidingen en een belangrijk deel van het schriftelijk onderwijs zijn daarin niet opgenomen. Om toch een schatting van het opleidingsniveau van de Nederlandse bevolking te kunnen maken wordt gebruik gemaakt van de EBB waarin informatie over ouderen en niet-bekostigd onderwijs wel is opgenomen. Bij de ouderen zijn de gegevens voornamelijk ontleend aan de EBB. In figuur 8 is de dekkingsgraad weergegeven van de gegevens uit het opleidingsregister. De groene lijn is de totale dekking, de zwarte lijn is de dekking op basis van de registraties en de rode lijn die van de EBB. De totale dekking is tot 14 jaar volledig, valt dan even sterk terug, om dan vanaf 18 jaar geleidelijk van 80% terug te lopen naar minder dan 10% voor de 55-plussers. Om vervolgens de gegevens een goede afspiegeling van de populatie te laten zijn is een weegmodel ontwikkeld voor verschillende peilmomenten (zie verder Bakker, Bouman en Van Toor, 2006; Bakker, Linder en Van Roon, 2008).

Figuur 6. De dekkingsgraad van het opleidingsregister



Het beroepsniveau is helaas niet in registers opgenomen. Daarom wordt die alleen in de enquête gemeten. Het *ln netto-uurloon* wordt als volgt gemeten. Eerst wordt het netto-jaarloon bepaald van de hoofd baan als het fiscaal loon verminderd met

de afgedragen premies en belastingen. Het fiscaal loon is ontleend aan de gegevens die bedrijven leveren aan de Belastingdienst voor het afdragen van belastingen en premies. Het fiscaal loon en de afgedragen premies en belastingen is benaderd voor de maand september 2004 door het jaarloon te delen door het aantal maanden dat de desbetreffende baan duurde die de respondent in september had. Daarmee is het een goede benadering van het begrip zoals dat in de enquête wordt gemeten. Het verschil betreft de bijzondere beloningen voor overwerk e.d. die in het netto-maandloon niet en in het fiscaal maandloon wel worden meegeteld. Voor het bepalen van het uurloon is gebruik gemaakt van de informatie zoals deze in de OSA-enquête voorkwam over het aantal gewerkte uren volgens contract. Dit was noodzakelijk omdat een goede registermeting voor het aantal contracturen voor het verslagjaar 2004 niet beschikbaar was.

### 3.3 *Koppeling van de data*

De koppeling van de data hebben we uitgevoerd door aan elk bestand een koppelnummer toe te kennen dat niet direct identificerend is voor een persoon of adres. Het OSA-aanbodpanel 2004 is gekoppeld met behulp van de sleutel geboortedatum, geslacht, postcode, en huisnummer. Dit levert een relatief hoog koppelerendement op van 98,9%. Van de 4 782 personen die deelnamen aan de enquête konden 4 730 personen van een koppelsleutel worden voorzien (Fouarge & Grim, 2007). Inspectie van de gegevens laat zien dat de uitgevallen 1,1% niet erg selectief is. Van deze 4 730 gekoppelde personen werken er 3 237 personen in loondienst.

Voor de registerinformatie geldt dat ook de afzonderlijke administratieve bestanden van een dergelijk koppelnummer zijn voorzien, waarbij vooral gebruik is gemaakt van het Sofinumnummer. De koppelerendementen van deze administratieve bestanden verschillen, maar ligt meestal boven de 97%. Bovendien wordt een belangrijk deel van de niet-gekoppelde records terecht niet gekoppeld, vanwege het feit dat deze niet tot de populatie behoren.

Vervolgens wordt gebruik gemaakt van het koppelnummer om het OSA-Aanbodpanel 2004 te koppelen met de informatie uit de registers. Daarbij kan nagenoeg iedereen met een koppelsleutel uit de OSA-enquête worden voorzien van een leeftijd en geslacht uit de registers. Ook is er nauwelijks sprake van gemiste koppelingen bij het uurloon. Dat ligt anders bij het opleidingsniveau.

Zoals eerder is gemeld wordt het opleidingsniveau vastgesteld door combinatie van gegevens uit registraties en dertien jaargangen EBB. Omdat de registraties vooral de jongere generaties en het bekostigde onderwijs in Nederland beschrijven, kunnen lang niet alle respondenten uit het OSA-Aanbodpanel van een opleidingsniveau worden voorzien. Van de 3 237 werknemers uit de OSA-enquête werken er 2 873 12 uur of meer. Hiervan kunnen er slechts 953 van een opleidingsniveau worden voorzien. Om toch de representativiteit te waarborgen is ervoor gekozen om de data te herwegen naar de verdeling over de combinatie van leeftijds-



klasse (in 10-jaarsklassen), geslacht en opleidingsniveau (5 categorieën) van 2 873 respondenten uit de OSA-enquête die meer dan 12 uur werken. De gewichten zijn zodanig bepaald dat ze optellen tot 953, dat wil zeggen dat ze een gemiddelde hebben van 1,0. Als er minder dan 3 respondenten in de 953 respondenten met een bekende opleiding in een cel voorkwam werd deze samengenomen met een naastgelegen cel. Dit kwam vooral bij ouderen voor. Verder is hetzelfde gedaan als er gewichten groter dan 5 voorkwamen. Steeds zijn dan leeftijdscategorieën samengenomen.

## 4. Resultaten

We starten eerst met het geven van de correlaties tussen de negen variabelen (Tabel 1). Daaruit kunnen we al veel afleiden. Leeftijd en geslacht worden vrijwel gelijk gemeten in de enquête en het register: de correlaties zijn nagenoeg 1.00. Het opleidingsniveau in het register meet wel iets anders dan dat in de enquête: de correlatie is slechts .80. Het ln netto uurloon heeft een iets grotere overeenkomst tussen het register en de enquêtes dan het opleidingsniveau. Ook hier geldt echter dat niet precies hetzelfde gemeten wordt: de correlatie is .83.

Tabel 1  
De correlaties tussen enquête- en registervariabelen

	leeftijd		geslacht		opleidingsniveau		beroepsniveau		
	enq	reg	enq	reg	enq	reg	enq	enq	reg
leeftijd uit enquête	1,000								
leeftijd uit register	0,998	1,000							
geslacht uit enquête	-0,097	-0,097	1,000						
geslacht uit register	-0,095	-0,096	0,996	1,000					
opleidingsniveau uit enquête	-0,072	-0,069	0,052	0,053	1,000				
opleidingsniveau uit register	-0,068	-0,066	0,075	0,070	0,797	1,000			
beroepsniveau uit enquête	0,083	0,084	-0,062	-0,063	0,566	0,584	1,000		
ln netto uurloon uit enquête	0,420	0,422	-0,121	-0,121	0,423	0,454	0,518	1,000	
ln netto uurloon uit register	0,477	0,479	-0,080	-0,081	0,391	0,404	0,495	0,827	1,000

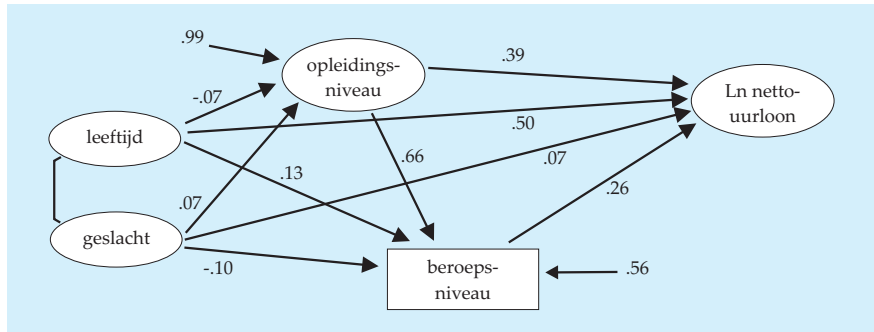
Het opleidingsniveau uit het register hangt hoger samen met het uurloon dan het opleidingsniveau uit de enquête. Daarbij maakt het niet uit of het uurloon uit het register of de enquête wordt genomen. Het opleidingsniveau uit het register hangt ook iets sterker samen met het beroepsniveau dan het opleidingsniveau uit de enquête. De enquêteversie van het uurloon hangt wel sterker samen met beroepsniveau dan de registerversie. De verschillen zijn echter zonder uitzondering vrij klein. Een voorlopige conclusie is dan ook dat de validiteit van de register- en enquêtevariabelen elkaar niet veel ontlopen.

De proef op de som wordt geleverd door de toepassing van het model zoals dat is weergegeven in figuur 5. Met behulp van LISREL 8 wordt de passendheid van dit

model berekend (Jöreskog & Sörböm, 1996; Kline, 2005). Het model past met een  $\chi^2$  van 69 bij 18 vrijheidsgraden. Er blijken geen residuele correlaties te zijn van enige omvang, dat wil zeggen dat de correlaties tussen de negen variabelen goed worden gereproduceerd door het model. Ook blijkt daarmee dat er geen sprake is van gecorreleerde meetfouten tussen de variabelen die in het model zijn opgenomen. Dat wil zeggen dat puur het feit dat variabelen uit een enquête of uit een register komt niet leidt tot een grotere samenhang tussen de meetfouten.

Vervolgens is van belang te beoordelen of het model plausible uitkomsten oplevert (figuur 7). Dat doet het. Het opleidingsniveau heeft een sterk positief effect op het beroepsniveau (.66) en een redelijk sterk direct effect op het uurloon (.39) los van het beroepsniveau. Daarnaast heeft opleidingsniveau nog een indirect effect op het uurloon via het beroepsniveau (.66 x .26 = .17). Het beroepsniveau heeft een positief effect op het uurloon, maar de correlatie tussen beroepsniveau en loon wordt voor een deel verklaard door een gemeenschappelijke samenhang met het opleidingsniveau.

**Figuur 7. Schatting van de validiteit van de registervariabelen leeftijd, geslacht, opleidingsniveau en in netto-uurloon**



Leeftijd heeft een sterk positief effect (.50) op het loon, los van het opleidings- en beroepsniveau en een eveneens positief effect op het beroepsniveau. Het effect is ook erg groot, groter dan verwacht en groter dan de correlaties tussen de metingen voor leeftijd en uurloon doen vermoeden. Blijkbaar is de correctie voor meetfouten die ervoor zorgt dat het effect van leeftijd wordt versterkt. Aangezien de meetfouten van leeftijd te verwaarlozen zijn, wordt dit veroorzaakt door de correctie voor meetfouten in het uurloon. Om dit te onderzoeken is het model toegepast op alleen op de OSA-data. Het effect van leeftijd is dan .40, wat redelijk in de buurt ligt van wat normaal wordt gevonden. Geslacht heeft kleine effecten op opleidingsniveau, beroepsniveau en uurloon. Vrouwen met een baan van meer dan 12 uur zijn hoger opgeleid dan hun mannelijke collega's, hebben een lager beroepsniveau en verdienen minder.

Uiteindelijk beoordelen we de omvang van de meetfouten (tabel 2). De meetfouten in leeftijd en geslacht zijn nihil en niet significant. De meetfouten in het opleidings-

niveau en uurloon zijn dat wel. Het opleidingsniveau in de enquête heeft een grotere meetfout dan dezelfde variabele uit het register. Voor het uurloon is het verschil in meetfout gering.

**Tabel 2**  
De meetfouten in de enquête- en registervariabelen

	enquête	register
leeftijd	0,00	0,00
geslacht	-0,01	0,01
opleidingsniveau	0,24**	0,17**
beroepsniveau	-	-
ln netto-uurloon	0,18**	0,17**

Significant  $p < .01$ .

## 5. *Conclusies en discussie*

Ondanks de toename van het gebruik van registratieve gegevens in sociaalwetenschappelijk onderzoek, is nog maar weinig bekend over mogelijke kwaliteitsproblemen ervan. In dit artikel wordt een methode gepresenteerd waarmee de validiteit vastgesteld kan worden van registervariabelen. Met behulp van de klassieke testtheorie en lineaire structurele modellen kan de validiteit van de registervariabelen worden vergeleken met enquêtevariabelen. Tevens kan de meetfout van de registervariabelen onder enkele aannames worden gekwantificeerd.

Dit wordt toegepast op een loonverwervingsmodel waarin leeftijd, geslacht, opleidingsniveau, beroepsniveau en uurloon zijn opgenomen. Met uitzondering van het beroepsniveau beschikken we over een meting uit een enquête en een uit een register. Uit de resultaten blijkt dat de metingen in enquêtes en registers elkaar weinig ontlopen in de omvang van de meetfouten. De uitzondering wordt gevormd door de variabele opleidingsniveau, die beter in het register is gemeten dan in de enquête.

Enkele kanttekeningen zijn op zijn plaats. Ten eerste betreft dit een eerste onderzoek waarin deze methode wordt toegepast. Het zou te vroeg zijn om de algemene conclusie te trekken dat registervariabelen meer valide worden gemeten dan enquêtevariabelen. Dit onderzoek zal meerdere keren moeten worden herhaald om tot meer generieke uitspraken te komen. Bovendien zal het beeld naar alle waarschijnlijkheid gemengd zijn. Sommige variabelen kun je heel goed meten in enquêtes met vraagstellingen die hun validiteit hebben bewezen, en sommige variabelen zijn veel lastiger in enquêtes te meten. Ook voor administratieve data geldt dat de ene variabele beter het concept meet dan andere.

Ten tweede is de meting van het opleidingsniveau aan de registerkant hybride: het betreft registermetingen aangevuld met enquêtegegevens. Dit nuanceert het beeld nog meer: de beste meting voor opleidingsniveau is een combinatie van registers en enquêtes. Het laat ook een nadeel van registers zien: soms ontbreekt een varia-

bele geheel of gedeeltelijk en dan moet je wel enquêteren om de variabele (bij) te schatten. Ten derde is het alleen toegestaan om de meetfouten te interpreteren als een maat voor de validiteit als met de latente variabelen inderdaad de ware scores worden gemeten. Dat wordt beoordeeld op grond van de parameters in het model waarmee de samenhangen tussen de latente variabelen worden beschreven. In ons voorbeeld is veel kennis over deze samenhangen, zowel van de theorieën als uit empirisch onderzoek. Dat is echter lang niet altijd het geval. De bruikbaarheid van de besproken methode wordt vergroot als deze kennis groter is.

## Referenties

Al, P. & B.F.M. Bakker (2000). Re-engineering Social Statistics by micro-integration of different sources. An introduction. In: P. Al & B.F.M. Bakker (red.), *Re-engineering social statistics by micro-integration of different sources*. Themanummer *Netherlands Official Statistics*, 15(summer) 4–6.

Arts, C.H. en E.M.J. Hoogteijling (2002). Het Sociaal Statistisch Bestand 1998 en 1999. *Sociaal Economische Maandstatistiek* 2002 (12), 66–71.

Bakker, B.F.M. (2008). De stand van het Sociaal Statistisch Bestand. *Bevolkingstrends*, 56 (2), 14–18.

Bakker, B.F.M. (2009a). *Micro-integratie* (Den Haag/ Heerlen: CBS).

Bakker, B.F.M. (2009b). *Trek alle registers open!* (Amsterdam: Vrije Universiteit).

Bakker, B.F.M., A.M. Bouman en L. van Toor (2006). Opleidingsniveau uit registers: nieuwe bronnen maar nog niet compleet, L. Engberts, F. Linder en F. Bastiaans (red.), *Sociale Samenhang in Beeld, het SSB nu en straks* (Voorburg/Heerlen: CBS), 141–162.

Bakker, B.F.M., F. Linder en D. van Roon (2008). Could that be true? Methodological issues when deriving educational attainment from administrative datasources and surveys (*Shanghai: Paper prepared for the IAOS Conference on Reshaping Official Statistics, 14–16 October 2008*).

Bakker, B.F.M. en L. Kuijvenhoven (2010). Registers en sociaalwetenschappelijk onderzoek: een geslaagde combinatie? B.F.M. Bakker en L. Kuijvenhoven, *Registers in sociaal-wetenschappelijk onderzoek. Mogelijkheden en valkuilen* (Den Haag / Heerlen: CBS).

Bethlehem, J. (2009). *Applied survey methods. A statistical perspective* (Hoboken New Jersey: Wiley)

Borghans, L. (2007). Anders denken, *Edata&research*, 2 (2), 8.

CBS (1993), *De Standaard beroepenclassificatie 1992* (Voorburg/Heerlen: CBS).

Fouarge, D. & R. Grim (2007). *Koppeling van het OSA-Arbeidsaanbodpanel aan administratieve gegevens: verslag en documentatie* (Tilburg: OSA).

Groves, R.M., F.J. Fowler jr., M.P. Couper, J.M. Lepkowski, E. Singer, en R. Tourangeau (2004). *Survey Methodology* (New York: Wiley Interscience).

Hart, H. 't, J. Van Dijk, M. de Goede, W. Jansen, en J. Teunissen (2005). *Onderzoeksmethoden* (Amsterdam: Boom)

Jöreskog, K., en D. Sörbom (1996). *LISREL: 8. User's reference guide* (Chicago: Scientific Software International).

Kline, R.B. (2005). *Psychological testing: a practical approach to design and evaluation* (New York: SAGE).

McCall, R.B. (2001). *Fundamental statistics for behavioural sciences* (Belmont: Wadsworth)

McQueen, R. en C. Knussen (2002). *Research methods for social science. An introduction* (Harlow: Prentice Hall)

NIWI (2010). *OSA Arbeidsmarktpanel 2000 voorlopige versie. P1462 Steinmetz Archive documentation set version 1.0* (Amsterdam: NIWI).

Singleton jr., R.A. en B.C. Straits (2005). *Approaches to social research* (Oxford / New York: Oxford University Press)

Statistics Denmark (1995). *Statistics on Persons in Denmark – A register-based statistical system* (Luxembourg: Eurostat).

Statistics Finland (2004). *Register based statistics. Best practices* (Helsinki: SF)

Wallgren, A. & B. Wallgren (2007). *Register-based statistics: administrative data for statistical purposes*. Wiley series in survey methodology (New York: Wiley)

---

<sup>1)</sup> De auteur is werkzaam als bijzonder hoogleraar Methodologie van registerdata aan de Vrije Universiteit van Amsterdam en als manager van de taakgroep Sociaal Economisch Totaalbeeld van het Centraal Bureau voor de Statistiek.



# *Participatie en loon na bedrijfseconomisch ontslag: Kapitaalvernietiging of negatief signaal?*

Didier Fouarge (ROA), Ruben van Gaalen (CBS) en Andries de Grip (ROA)

## *1. Inleiding*<sup>1</sup>

De centrale vraagstelling in dit hoofdstuk is wat de gevolgen zijn van bedrijfseconomisch ontslag op de participatiekans en het loon na het ontslag. De analyses zijn uitgevoerd met het Sociaal Statistisch Bestand (SSB) van het CBS, dat recentelijk is uitgebreid met informatie over ontslag om bedrijfseconomische redenen. De administratieve data omvatten een grote populatie ontslagenen om bedrijfseconomische reden, waarin het mogelijk is om onderscheid te maken naar drie soorten ontslag: ontslag vanwege faillissement, UWV-ontslagaanvragen en collectieve uitstroom. Ontslag en werkloosheid zijn onderwerpen waarover mensen in enquêtes minder geneigd zijn te antwoorden. Bovendien geven ze sneller sociaal wenselijke antwoorden. De administratieve werkelijkheid daarentegen is niet afhankelijk van selectieve respons noch van het eventueel selectieve geheugen van respondenten. Uit onderzoek blijkt dat werkloosheid een blijvend negatief effect heeft op het loon later in de loopbaan. In deze bijdrage gaan wij na of dit het gevolg is van depreciatie van menselijk kapitaal tijdens de werkloosheid. Vanuit het perspectief van de theorie van het menselijk kapitaal gaat werkloosheid inderdaad gepaard met een periode waarin het menselijk kapitaal niet gebruikt wordt, waardoor kennis en vaardigheden aan waarde inboeten. wat vervolgens resulteert in een lagere kans op werk en een lager loon na werkloosheid (Pissarides, 1992). Een alternatieve verklaring hiervoor is echter dat er een signaalfunctie uitgaat van werkloosheid (Albrechts e.a., 1999): de kans op werkloosheid is selectief waardoor de minst productieve werknemers meer kans hebben om werkloos te worden. Potentiële nieuwe werkgevers reageren hierop door deze werknemers een lager loon te bieden dan werknemers met evenveel ervaring die niet werkloos zijn geweest.

Dat de populatie ontslagenen slechtere arbeidsmarktkansen zou hebben komt doordat deze werknemers simpelweg minder geschikt zijn en niet door de gedwongen loopbaanonderbreking op zich. Dit probleem van selectiviteit in de kans op werkloosheid is in de literatuur methodologisch 'opgelost' door gebruik van de 'control function' methode (Heckman en Robb, 1985) of instrumentele variabele (Gregg en Tominey, 2005). Daarbij probeert men te controleren voor de selectiviteit

in de kans op werkloosheid. Het probleem hierbij is dat men afhankelijk is van de aanwezigheid van goede instrumenten in de data die de kans op werkloosheid verklaren maar niet samenhangen met de uitkomstvariabele zoals de latere participatiekans en het latere loon. Hierbij blijft het onzeker of de gebruikte variabelen voor het identificeren van zelfselectie hun werk inderdaad goed doen.

De 'state of the art' in de economische literatuur is dan ook om gebruik te maken van data rond 'displaced workers' (ontslagenen door massaontslag of faillissementen) om het echte effect van werkloosheid op de participatiekans en het loon te schatten (Kletzer, 1998; Kriechel, 2003; Koeber en Wright, 2006; Song, 2009). De reden hiervoor is dat bij collectief ontslag geen sprake is van eigen schuld van de ontslagene (Kriechel, 2010), en deze vorm van werkloosheid veel weg heeft van een natuurlijk experiment waarin iemand 'bij toeval' werkloos is geworden. Hierdoor ziet de nieuwe werkgever de werkloosheid niet als signaal voor een lage productiviteit.

De vraag die we willen beantwoorden is: wat zijn de gevolgen van ontslagwerkloosheid voor de participatiekans en het loon na werkloosheid, rekening houdend met de selectiviteit van de groep ontslagenen? Eerdere Nederlandse studies hebben zich moeten beperken tot het faillissement van één bedrijf (Fokker in 1996; zie Kriechel, 2003; Kriechel en Pfann, 2005), collectieve uitstroom (Scheele, van Gaalen, & van Rooijen, 2008), of collectieve uitstroom en UWV ontslagaanvragen (Van Gaalen en Van Rooijen, 2009). Onze analyses zijn uitgevoerd met het Sociaal Statistisch Bestand (SSB) van het CBS, dat recentelijk is uitgebreid met informatie over ontslag om bedrijfseconomische redenen.

## 2. *Eerder onderzoek*

### 2.1 *Werkloosheid en selectie*

Bezien vanuit de menselijk kapitaal theorie is de relatieve depreciatie van menselijk kapitaal de oorzaak van de lagere baankans en het lagere loon van werkenden na een periode van inactiviteit (Pissarides, 1992). Niet alleen doen werklozen tijdens de periode van werkloosheid geen werkervaring op en is er dus sprake van niet-opbouw van menselijk kapitaal, ook is de bedrijfsspecifieke kennis die iemand vóór de onderbreking door werkloosheid heeft opgedaan veelal onbruikbaar in de nieuwe baan. Hier is sprake van depreciatie: het menselijk kapitaal is minder waard geworden. Uit de empirische studies naar deze 'human capital hypothese' komt naar voren dat kennis en vaardigheden afnemen tijdens een periode van niet-werkzaamheid, in het bijzonder tijdens werkloosheid (Edin en Gustavson, 2008) en de zorg voor kinderen (Albrechts e.a., 1999).

De signaleringstheorie stelt echter dat het niet zozeer gaat om de depreciatie van kennis en vaardigheden, maar eerder om de signaalfunctie die uitgaat van werkloosheid en andere soorten van inactiviteit (Albrechts e.a., 1999). Omdat de kans op werkloosheid selectief is, zijn het vooral de minst productieve werknemers die kans



hebben om werkloos te worden. Potentiële nieuwe werkgevers herkennen dit signaal en zullen deze werknemers een lager loon bieden dan werknemers met evenveel ervaring die niet werkloos zijn geweest. Daarbij moet wel opgemerkt worden dat studies veelal laten zien dat werklozen deze loonachterstand voor een deel weer inlopen. Bijvoorbeeld, voor Nederland laten Schils e.a. (2006) aan de hand van het *OSA-Arbeidsaanbodpanel* zien dat een periode van werkloosheid een nadelig effect heeft op de baankans later in de loopbaan, maar dat dit negatieve effect na ongeveer zes jaar helemaal verdwijnt. Verder blijkt dat werkloosheid wel een blijvend negatief effect heeft op het loon later in de loopbaan: de loonachterstand ten gevolge van werkloosheid wordt nooit meer ingehaald. Dit effect van werkloosheid – ook wel het ‘scar’ effect van werkloosheid genoemd (Gangl, 2006; Mooi-Reçi, 2008) – wijst er op dat werkgevers zich bij het in dienst nemen van een werkloze indekken tegen de risico’s. Als ze na enige tijd betere informatie hebben over de productiviteit van hun nieuwe medewerker, dan zijn ze bereid een passend loon aan te bieden.

In empirisch onderzoek is het van belang om de signaalfunctie van werkloosheid te scheiden van de depreciatie van het menselijk kapitaal. Dit kan door het onderzoeksdesign zorgvuldig te kiezen of door rekening te houden met mogelijke selectie-effecten van loopbaanonderbrekingen. Om het depreciatieeffect van werkloosheid te kunnen vaststellen maken sommige studies bijvoorbeeld gebruik van natuurlijke experimenten die kunnen ontstaan bij veranderingen in de ww-regelingen (Mooi-Reçi, 2008). Andere studies maken gebruik van ‘matching’ technieken of van de ‘control function’ aanpak (Schils e.a., 2006). Een aantal studies lijkt bewijs voor deze signaalhypothese te vinden doordat de omvang van de loondaling in de nieuwe baan samenhangt met de reden voor werkloosheid: mensen die werkloos zijn geworden door een faillissement ervaren een minder sterke loondaling dan mensen die ontslagen zijn wegens disfunctioneren (zie bijvoorbeeld Gibbons en Katz, 1991).

## 2.2 *Bedrijfseconomisch ontslag*

Er bestaat een uitgebreide literatuur waarin specifiek ingegaan wordt op het verlies van bedrijfsspecifieke kennis als gevolg van werkloosheid door massaontslag. Deze literatuur over ‘displaced workers’ wordt vooral gevoed door Amerikaanse studies (zie bijvoorbeeld het overzicht van Scheele et al., 2008). Europa, en in het bijzonder Nederland, loopt op dit terrein enigszins achter. Het voordeel van deze focus op ‘displaced workers’ is dat werkloosheid ten gevolge van massaontslag of faillissementen niet te wijten is aan het disfunctioneren of andere tekortkomingen van de ontslagen werknemer zelf. Een dergelijke setting maakt het dus mogelijk om het effect van werkloosheid op de waarde van iemands menselijk kapitaal te bepalen, zonder dat er een vertekening optreedt vanwege een selectie-effect (Kriechel, 2010).

Hamermesh (1987) liet op basis van Amerikaanse data zien dat ‘displaced workers’ een aanzienlijk deel van hun bedrijfsspecifieke menselijk kapitaal verliezen. Dit komt tot uitdrukking in een lager loon na werkloosheid. Een belangrijk onderzoek

op dit terrein is dat van Jacobson e.a. (1993). Zij gebruiken administratieve data voor de Amerikaanse staat Pennsylvania om de effecten van werkloosheid als gevolg van massaontslag te kwantificeren. Zij vinden grote effecten van een periode van werkloosheid op het loon bij herintrede; het loonverlies bedraagt maar liefst 25% van het oude loon. Daarbij is het inkomensverlies van mannen groter dan het inkomensverlies van vrouwen. Bovendien laten de auteurs ook zien dat het loon al in de vier jaar voor het massaontslag lager wordt. Hierdoor is er mogelijk al voor het massaontslag sprake van een selectie-effect doordat de werknemers met de betere externe arbeidsmarktperspectieven overstappen naar een ander bedrijf. Echter, de analyses van Hammermesh (1987) wijzen niet op een dergelijke selectie vóór het massaontslag.

Kletzer (1998) laat op basis van het Amerikaanse *Current Population Survey* zien dat het effect van werkloosheid op het loon afhankelijk is van de mate waarin het bedrijfsspecifieke menselijk kapitaal overgedragen kan worden van de ene naar de andere baan. Zo blijkt het inkomensverlies van werklozen die van sector veranderen groter te zijn dan dat van werknemers die een nieuwe baan hebben gevonden binnen dezelfde sector. Een verklaring hiervoor is dat werknemers door deze sectorverandering hun sectorspecifieke menselijk kapitaal niet meer kunnen inzetten. Volgens het onderzoek van Neal (1995) is het inkomensverlies als gevolg van massaontslag het grootst voor werknemers met veel dienstjaren die van sector veranderen.

De analyses door Burda en Mertens (2001) laten ook een inkomensverlies na massaontslag zien in Duitsland op basis van de *German Socio-economic Panel* (GSOEP) en data van het Institut für Arbeitsmarkt- und Berufsforschung (IAB). De inkomensverliezen zijn hier eveneens groter voor werknemers die na ontslag van sector zijn veranderd, maar ook groter voor werknemers aan de bovenkant van de loonverdeling. Eveneens op basis van de GSOEP laat Couch (2001) zien dat werklozen als gevolg van massaontslag weliswaar een lager loon verdienen in hun nieuwe baan, maar dat de inkomensachterstand na verloop van tijd deels wordt ingehaald. Het gemeten effect op het loon in Duitsland is ongeveer vergelijkbaar met de laagste effecten die voor Amerika worden gevonden. In termen van de kans op het vinden van nieuw werk laat de studie van Couch zien dat deze aanvankelijk lager is voor personen die recentelijk werkloos waren, maar dat het na een paar jaar niet meer uitmaakt of men wel of niet eerder werkloos is geweest.

Voor het Verenigd Koninkrijk vinden Gregory en Jukes (2001) op basis van de data van het *New Earnings Survey Panel* blijvende negatieve effecten van werkloosheid op het loon. Deze effecten, zijn, zoals ook in ander onderzoek is aangetoond, groter voor ouderen (hoewel voor hen de kans op massaontslag kleiner is) en hoogopgeleiden. Arulampalam (2001) vindt op basis van de *British Household Panel Study* negatieve en blijvende inkomenseffecten van werkloosheid. Zij laat bovendien zien dat dit effect kleiner is voor werklozen die om bedrijfseconomische redenen zijn ontslagen dan voor werklozen die om andere redenen zijn ontslagen. Meer recent is de studie van Hijzen e.a. (2010) op basis van verschillende databronnen voor het Verenigd Koninkrijk. Hun conclusie is dat een periode van werkloosheid een negatief effect heeft op ie-

mand's loon, maar dat dit effect kleiner is voor werknemers die werkloos zijn geworden door massaontslag dan voor werklozen als gevolg van een faillissement<sup>2</sup>. In hun studie verdwijnt het negatieve looneffect van werkloosheid na ongeveer vijf jaar.

In Nederland is er minder onderzoek naar het verlies van menselijk kapitaal bij ontslagwerklozen Abbring e.a. (2002) gebruiken enquêtegegevens uit het *OSA-Arbeidsaanbodpanel* en maken een vergelijking met de VS. Zij vinden voor Nederland langere werkloosheidsduren. Wel blijkt werkloosheid als gevolg van massaontslag de uitstroom naar vervroegde uittrekking te stimuleren. Ook vinden ze dat in Nederland het loon in de nieuwe baan na ontslag hoger is dan het loon in de oude baan. Dit suggereert dat een vergelijking van de loongroei van ontslagen en niet-ontslagen werknemers economisch relevanter is dan de vergelijking van het loonniveau voor en na het ontslag. Een mogelijke verklaring voor het ontbreken van een negatief effect op het loon is dat werknemers op een massaontslag anticiperen en vooraf al op zoek gaan naar een andere baan. Verder bestaat het Nederlands onderzoek op dit terrein vooral uit analyses van de gevolgen van het faillissement van Fokker in 1996 (Kriechel, 2003). Hoewel de ontslagen Fokker werknemers betrekkelijk snel na het faillissement weer een baan hebben gevonden (77% had na een jaar weer werk, na drie jaar was dat 85%) zijn de geleden inkomensverliezen gevoelig: per maand van werkloosheid bedraagt het loonverlies 0,65%. De participatiekans na het faillissement is kleiner voor oudere en laagopgeleide werknemers. Bovendien is het loonverlies voor oudere werknemers significant groter (Kriechel en Pfann, 2005). Ook Bakker et al. (1999) laten op basis van het Socio-Economisch Panel (SEP) zien dat werkloosheid alleen bij mannen bestraft wordt met een lager toekomstig loon<sup>3</sup>.

Meer recent hebben de WRR en het CBS een studie gepubliceerd over massaontslag in Nederland (Scheele et al., 2008). De studie is gebaseerd op analyses van het *Sociaal Statistisch Bestand (SSB)*. Massaontslag is in de studie gedefinieerd als een uitstroom van werknemers uit een bedrijf die hoger ligt dan de normale uitstroom uit het bedrijf. Geconcludeerd wordt dat tweederde van de werknemers een half jaar na ontslag weer werk heeft; 3 jaar na ontslag is dit 68%. Wat het inkomen betreft concludeert de WRR opmerkelijk genoeg dat massaontslag een gunstige dynamiek tot stand weet te brengen doordat werkloosheid een positief effect op het loon heeft. Hier zou sprake kunnen zijn van heterogeniteit in het effect van werkloosheid op het loon na werkloosheid waarbij het vooral jongeren zijn die beter af zijn terwijl ouderen juist terecht komen in banen met een lager loon.

### 3. *Data: Sociaal Statistisch Bestand (SSB)*

We gebruiken registratiegegevens van de uitstroom van werknemers uit bedrijven zoals die opgenomen zijn in het Sociaal Statistisch Bestand (SSB) van het CBS. Het SSB is, anders dan de naam doet vermoeden, niet één enkel bestand. De definitieve versie ervan is een stelsel van koppelbare registers en enquêtes die onderling op

elkaar zijn afgestemd en consistent zijn gemaakt (Arts & Hoogteijling, 2002; Bakker, 2008, 2009). Per jaargang worden inmiddels bijna 60 registers gebruikt. Daarnaast zijn ook gegevens uit enkele enquêtes in het SSB opgenomen. De demografische gegevens van het SSB (geslacht, geboortedatum, geboorteland, burgerlijke staat en woonplaats etc) komen uit de Gemeentelijke basisadministratie persoonsgegevens (GBA).

Wat betreft bedrijfseconomisch ontslag maken we in onze analyse gebruik van de recent beschikbaar gekomen SSB-gegevens over bedrijfseconomisch ontslag. In Nederland wordt het grootste deel van de onvrijwillige mobiliteit van rechtswege beëindigd (bijvoorbeeld een aflopend tijdelijk contract; in 2006 76%, zie Knegt en Tros, 2007). In ons onderzoek onderscheiden we drie vormen van bedrijfseconomisch ontslag in 2005: 1) Ingewilligde individuele ontslaaanvragen wegens bedrijfseconomische redenen bij het UWV werkbedrijf (N=35 000). 2) In geval van een faillissement is geen ontslagvergunning benodigd. Faillissementen worden ontleend aan rechtbankregistraties. Baanbeëindigingen in het uitspraakjaar of het jaar daarvoor worden ook tot faillissementsontslag gerekend (N=31 000). 3) Collectieve ontslagen: In grote bedrijven worden ontslagrondes vaak afgewikkeld via een sociaal plan (ontslag met wederzijds goedvinden), zonder dat er een instantie aan te pas komt (Scheele, Theeuwes, & de Vries, 2007). Piekuitstroom in een bepaalde maand wordt afgeleid uit het SSB (N=14 000). Deze drie vormen van bedrijfseconomisch ontslag vertegenwoordigt ongeveer 50% van alle vaste dienstverbanden die om bedrijfseconomische redenen zijn beëindigd (Knegt en Tros, 2007). Ontslag op staande voet wordt bijvoorbeeld niet geregistreerd. Vaak vindt ontslag met wederzijds goedvinden geleidelijk over de tijd verdeeld plaats. Ook in kleine bedrijven worden veel banen met wederzijds goedvinden beëindigd. Een deel (ontbindingsverzoeken) wordt ook afgehandeld door het Kantongerecht (zie ook Bloemendal, 2010; Corpeleijn, 2009; Van Gaalen & Van Rooijen, 2009; SZW 2000, 2009).

Om te bepalen in hoeverre ontslagenen om bedrijfseconomische redenen verschillen van niet-ontslagenen wordt een vergelijkingsgroep geformeerd uit de populatie werknemers (ruim 6 miljoen personen) die het gehele jaar werknemer waren. Voor de analyses hebben we per persoon één record per maand in de periode 2003–2007, om op deze wijze iedere verandering in loon en participatie in beeld te brengen. Om de hoeveelheid data hanteerbaar te maken trekken we een gerandomiseerde steekproef van 10 000 werknemers: 6 083 niet-ontslagenen, 1 879 ontslagenen via het UWV (bedrijfseconomisch), 1 378 faillissementsontslagenen, en 660 collectief ontslagenen. De ontslagen werknemers worden gevolgd tot 2008 en vergeleken met niet-ontslagen werknemers.

De eerste afhankelijke variabele, participatie voor en na ontslag, is het wel of niet hebben van een baan als werknemer. De tweede afhankelijke variabele, het loon voor en na ontslag, is het bruto maandloon van een baan als werknemer, omgerekend naar een volledige dienstbetrekking. Het ijkpunt voor de ontslagenen is de ontslagmaand; voor de niet-ontslagenen is dit een willekeurige kalendermaand in 2005.

## 4. *Beschrijvende resultaten*

### 4.1 *Relatie persoonskenmerken en ontslag*

Op basis van de verdelingen naar leeftijd en geslacht lijken de UWV ontslagenen het meest op de niet-ontslagenen (tabellen 1–3). Wat opvalt is het hoge percentage jongeren onder de faillissementontslagenen<sup>4</sup> en de relatief grote aandeel ouderen én jongeren in de groep collectief ontslagenen. Van de groep niet-ontslagen personen die geheel 2005 werkzaam waren is 16% autochtoon. Het percentage allochtonen is binnen de groep ontslagen werknemers duidelijker hoger: gemiddeld 25%. De groep overig niet-westerse allochtonen wordt bovengemiddeld vaak met ontslag geconfronteerd. Dit verschil is deels terug te voeren op faillissementsontslag, waar deze groep allochtonen relatief vaak mee te maken heeft.

### 4.2 *Relatie baan- en bedrijfskenmerken en ontslag*

De gemiddelde baanomvang op tijdstip  $t=0$  voor de faillissementsontslagenen is relatief laag. Dit betreffen vaker dan gemiddeld kleine baantjes. De gemiddelde baanduur (niet in tabel) laat zien dat het met name kortdurende baantjes zijn. Opvallend is dat ongeveer 5% van de personen wiens baan werd beëindigd voordat het desbetreffende bedrijf failliet ging geen baan heeft. Dit heeft te maken met het feit dat de maandgegevens peildatumgegevens zijn, in dit geval op de derde vrijdag van iedere maand. Juist bij deze groep werknemers komt het voor dat de baan nog niet bestond op peildatum [ $t=-1$ ] (één maand voor het ontslag) en niet méér bestond op  $t=0$ . De collectief ontslagenen hebben het vaakst een betrekking in vol-tijd en voor hen is de baanduur relatief gezien het langst (gemiddeld ruim 8 jaar). Wat betreft de bedrijfskenmerken valt op dat er in 2005 vooral banen binnen de zakelijke dienstverlening verloren gingen als gevolg van een faillissement (zie tabel 3). Dit gebeurt veel minder vaak in gezondheid- en welzijnsinstellingen. In de industrie en de sector handel/repairatie gaan relatief veel banen verloren die via de UWV route lopen. Collectief ontslag c.q. sociale plannen komen vaker voor in de publieke sector (Openbaar bestuur en Onderwijs).

Het is niet verassend dat de collectief ontslagenen voornamelijk in hele grote bedrijven werkzaam waren (tabel 2); dit was deel van het selectieproces voor de constructie van de data (zie paragraaf 3). Opvallend is wel dat het algemene bedrijvenregister (ABR), waar dit kenmerk op is gebaseerd, enige afwijking vertoont met het gemiddelde aantal banen dat in het SSB wordt waargenomen: een klein percentage van de collectief ontslagenen zou volgens de ABR in kleine bedrijven werkzaam zijn geweest. Administratieve vertraging is één mogelijke oorzaak van deze discrepantie. Het ABR geeft een jaarcijfer, terwijl in geval van een massaontslag het personeelsbestand van een bedrijf gedurende een jaar juist sterk wordt verkleind. Verder zien we dat bedrijven die failliet gaan vaak erg klein zijn. Personen die via het UWV worden ontslagen komen zowel uit grote als kleine, maar ook vaker dan gemiddeld uit kleine bedrijven.

**Tabel 1**  
**Persoonskenmerken (niet-)ontslagenen**

	Geen ontslag	Faillissement	UWV	Collectief
	%			
<i>Leeftijd</i>				
34 jaar en jonger	36	55	33	50
35-39 jaar	14	11	16	7
40-44 jaar	14	10	14	8
45-49 jaar	12	8	13	6
50-54 jaar	11	7	12	7
55-59 jaar	10	6	10	13
60 jaar en ouder	3	3	3	9
<i>Geslacht</i>				
Mannen	55	67	56	61
Vrouwen	45	33	44	39
<i>Herkomstgroepering</i>				
Autochtonen	84	74	81	83
Marokkanen/Turken	3	7	4	3
Surinamers/Antillianen	3	4	2	2
Overig Niet-Westers	2	5	4	3
Westers	8	11	9	9

Bron: SSB.

**Tabel 2**  
**Baankenmerken (niet-)ontslagenen**

	Geen ontslag	Faillissement	UWV	Collectief
	%			
<i>Baanomvang (maand=0)</i>				
Geen baan	0	6	0	0
0-20%	3	6	3	2
20-40%	7	13	6	5
40-60%	11	9	10	9
60-80%	10	11	10	12
80-100%	69	55	70	71
<i>Fiscaal loon (maand=0)</i>				
Geen ink	0	6	0	0
≤1 000	6	20	7	17
1 001-2 000	19	26	26	23
2 001-3 000	40	29	32	19
3 001-4 000	19	10	14	17
4 001-5 000	7	5	7	9
≥ 5 000	8	5	14	16
<i>Grootteklasse (ABR)</i>				
Onbekend	1	6	1	0
<10 werknemers	14	34	34	1
10- 49 werknemers	17	30	24	2
50- 99 werknemers	8	12	7	1
100-499 werknemers	19	12	14	12
≥ 500 werknemers	41	6	20	84

Bron: SSB.

Tabel 3  
Bedrijfssectoren (niet-)ontslagenen

	Geen ontslag	Faillissement	UWV	Collectief
	%			
Landbouw, bosbouw, visserij	1,6	1,0	2,6	0,0
Delfstoffenwinning	0,2	0,1	0,0	0,3
Industrie	13,4	13,1	20,8	7,7
Energie- en waterschappen	0,4	0,2	0,1	2,1
Bouwnijverheid	5,1	7,7	9,1	0,5
Handel en reparatie	16,0	15,0	25,4	22,6
Horeca	2,9	6,2	3,9	0,6
Vervoer, opslag	5,9	10,1	5,3	12,9
Financiële instellingen	4,0	3,1	1,0	0,0
Zakelijke dienstverlening	14,7	27,1	16,0	7,4
Openbaar bestuur	8,2	0,4	0,4	14,2
Onderwijs	6,8	2,0	0,9	20,8
Gezondheids- en welzijn	16,1	5,5	10,5	7,3
Milieu, cultuur	3,7	2,8	3,7	3,5
Onbekend	1,0	5,8	0,2	0,2
Totaal	100	100	100	100

Bron: SSB.

## 5. *Gevolgen ontslag voor participatie en loon*

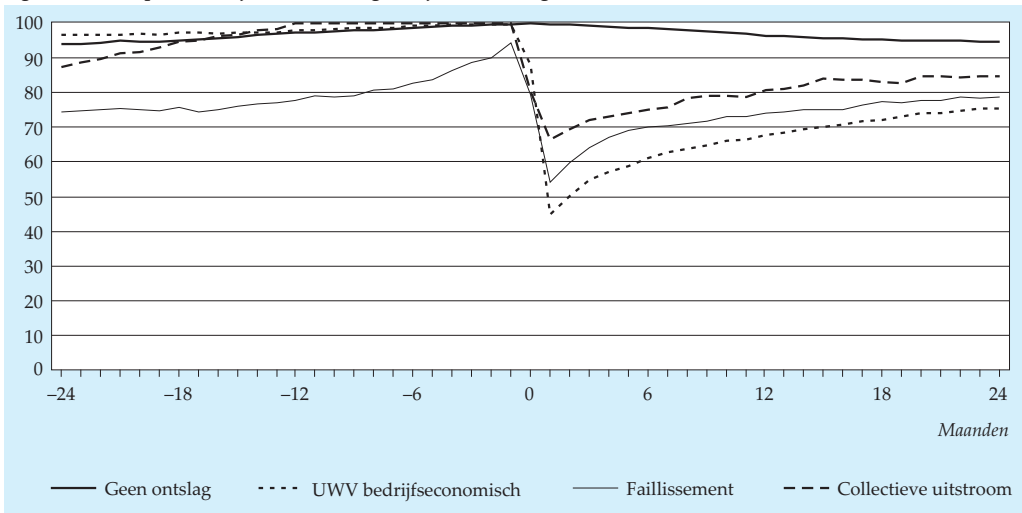
### 5.1 *Participatie en loon voor en na ontslag*

In tegenstelling tot de eerder gepresenteerde cijfers beperken we ons vanaf nu tot degenen die na 1945 zijn geboren en sluiten wij de personen die na ontslag met pensioen gaan uit. In figuur 1a wordt per maand de gemiddelde participatie en in figuur 1b wordt het fiscaal maandloon van 2003–2007 in beeld gebracht. In het bijzonder onder de werknemers die collectief stoppen met werken bevinden zich veel werknemers die met (pre)pensioen gaan (18%). De maandloonlijnen worden alleen berekend voor de mensen die daadwerkelijk een baan hebben.

Een groot deel van de ontslagenen vindt direct aansluitend nieuw werk: op het moment van ontslag zakt de participatie scherp maar niet verder dan 55% (faillissement), 45% (UWV) en 65% (collectief). Vervolgens stijgt het participatieniveau, maar het blijft om en nabij de 80% steken. Personen die met een collectieve regeling stoppen en niet met (pre)pensioen gaan doen het iets beter dan de andere ontslagenen.

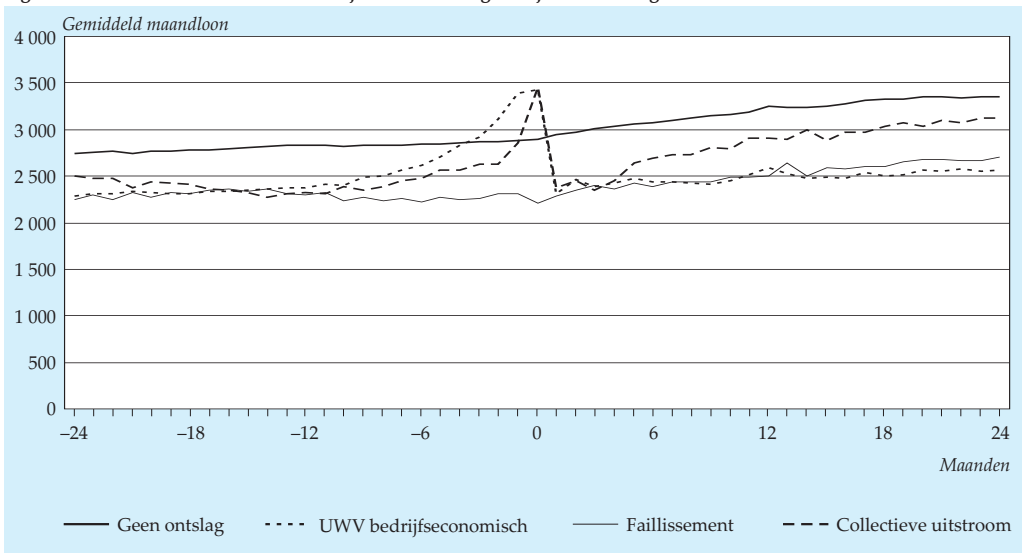
Interessant is dat de lijn bij de faillissementsontslagenen licht stijgt vóór  $t=0$ . Dit hangt samen met de al genoemde selectiviteit van relatief veel korte baantjes. Ook zien we dat de lijn de 100% niet haalt: dit heeft ook te maken met de eerdergenoemde peildatummeting. Het fiscale loon neemt bij de niet-ontslagenen geleidelijk toe. Het gemiddelde salaris bij de ontslagenen ligt gedurende de gehele periode lager dan dat van de niet-ontslagenen. Verder stijgt het niet of in elk geval niet zo snel als in de vergelijkingsgroep. De sterke stijging van het loon vóór ontslag (behalve bij de faillissementsontslagenen) is opmerkelijk. In dit kader is het van belang dat in het SSB het betalingsprincipe wordt gehanteerd. Een baan bestaat als er loon wordt betaald. Het

**Figuur 1a. Participatie van 2 jaar voor ontslag tot 2 jaar na ontslag**



Bron: SSB.

**Figuur 1b. Gemiddeld maandloon van 2 jaar voor ontslag tot 2 jaar na ontslag**



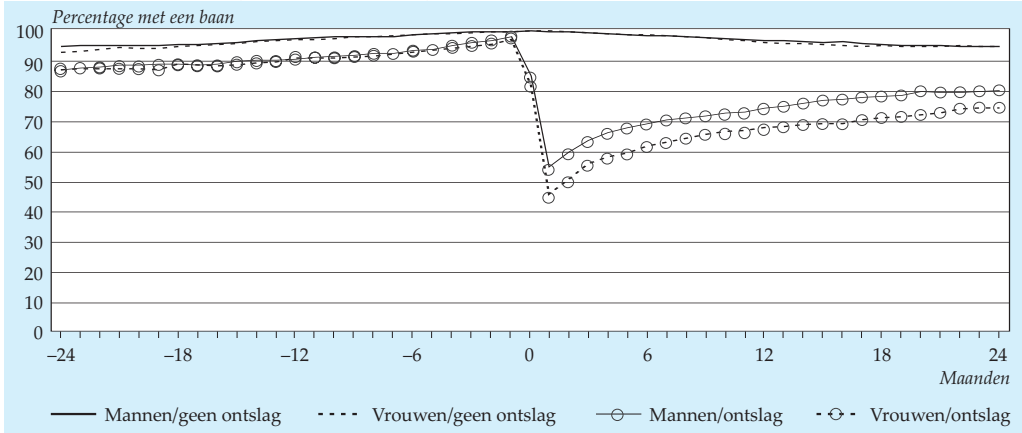
Bron: SSB.

in een jaar betaalde loon wordt over de maanden dat men in dienst was verdeeld. Daarmee samenhangend heeft de op het oog wat vreemde loonpiek te maken met het feit dat mensen ontslagvergoedingen krijgen. Deze worden over minder kalendermaanden verdeeld wat leidt tot een hoger gemiddeld maandloon.



Het effect van ontslag op de arbeidsparticipatie van mannen is kleiner dan het effect op de arbeidsparticipatie van vrouwen (figuur 2a–b). Het effect van ontslag op het loon doet zich voor bij zowel mannen als vrouwen: de loonontwikkeling blijft achter vergeleken met die van werknemers die niet zijn ontslagen. Overigens blijkt dat jongeren vooral wat betreft de baankans na ontslag betere kaarten hebben dan ouderen (niet getoond in figuur).

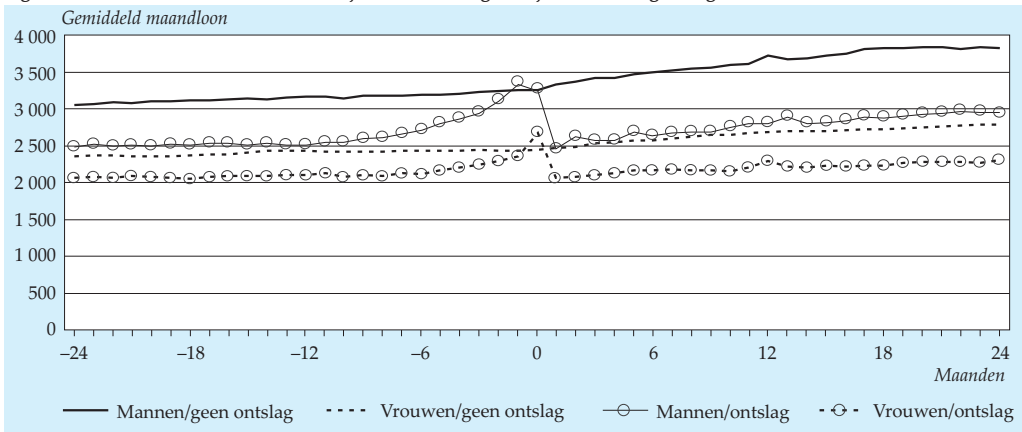
Figuur 2a. Participatie van 2 jaar voor ontslag tot 2 jaar na ontslag naar geslacht<sup>1)</sup>



<sup>1)</sup> In de figuur zijn de drie vormen van bedrijfseconomisch ontslag bij elkaar genomen.

Bron: SSB.

Figuur 2b. Gemiddeld maandloon van 2 jaar voor ontslag tot 2 jaar na ontslag naar geslacht



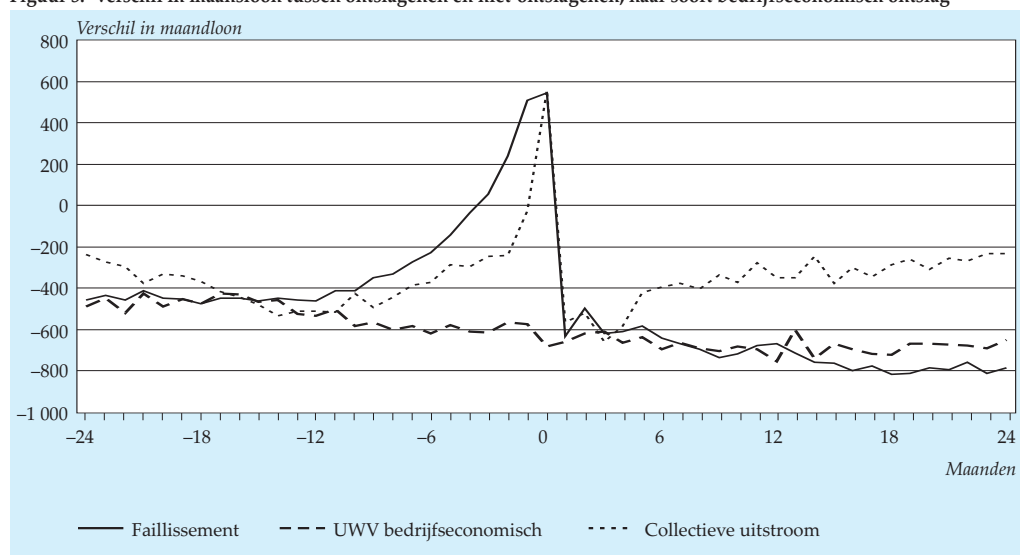
<sup>1)</sup> In de figuur zijn de drie vormen van bedrijfseconomisch ontslag bij elkaar genomen.

Bron: SSB.

Figuur 3 vergelijkt de loonontwikkeling van ontslagwerklozen voor en na het ontslag met de loonontwikkeling van niet-werklozen. Het is duidelijk dat het loon van ontslagwerklozen lager is dan het loon van de referentiegroep in de maanden voor ontslag. Alleen voor de collectieve uitstroom en voor UWV ontslagwerklozen neemt het loon toe in de paar maanden voor ontslag, in overeenstemming met de ontwikkeling in figuur 1. Hierdoor wordt voor de ontslagwerklozen in de maanden vlak voor het ontslag een hoger loon gemeten dan voor niet ontslagenen. Zoals eerder gezegd heeft dit met ontslagvergoedingen te maken.

Al met al kan op basis van de data worden vastgesteld dat het loonverschil tussen ontslagenen en niet-ontslagen groter is na het ontslag. Dit geldt althans voor werknemers die een bedrijfseconomisch ontslag via de UWV of een faillissement hebben meegemaakt. Voor de collectieve uitstroom geldt dat ontslagenen in de eerste 24 maanden na ontslag hun loonachterstand op niet-ontslagenen deels inlopen.

Figuur 3. Verschil in maansloon tussen ontslagenen en niet-ontslagenen, naar soort bedrijfseconomisch ontslag



Bron: SSB.

## 5.2 Multivariate analyses

De kans op een baan en het loon na ontslag zijn natuurlijk niet alleen afhankelijk van het soort bedrijfseconomisch ontslag. Zo laten de gegevens in paragraaf 4 zien dat ontslagenen en niet-ontslagenen op enkele achtergrondkenmerken verschillend zijn van elkaar. Dit geldt ook voor de verschillende vormen van bedrijfseconomisch ontslag onderling. Om die reden is het van belang om in multivariate analyses te controleren voor deze achtergrondkenmerken.<sup>5</sup> Daarbij onderscheiden wij drie modellen. Ten eerste kijken wij naar de kans op participatie een jaar na het bedrijfseconomisch ontslag en

vergelijken wij deze kans met de participatiekans voor niet-ontslagenen. Ten tweede kijken wij naar het loon 12 maanden na ontslag. Ten slotte maken wij gebruik van een selectiemodel om het effect van bedrijfseconomisch ontslag op het loon 12 maanden na ontslag te bepalen, conditioneel op het feit dat de betreffende persoon weer een baan heeft gevonden. Model 1 in tabel 4 rapporteert de schattingsresultaten van een logit model met betrekking tot de kans op participatie 12 maanden na ontslag. De gepresenteerde cijfers zijn de odds-ratios<sup>6</sup>. De bevindingen uit de beschrijvende analyses worden bevestigd. De baankans na 12 maanden is significant lager voor ontslagenen dan voor niet-ontslagenen. De controlevariabelen wijzen de verwachte kant op: jongeren, mannen, en autochtonen hebben vaker een baan. In het model wordt gecontroleerd voor de werkgelegenheidsontwikkeling voor de sector waarin men werkzaam was in de maand van ontslag. Dit is gedaan aan de hand van ROA prognoses voor de werkgelegenheidsontwikkeling (ROA, 2009). Werkt(e) iemand in een groeisector op tijdstip  $t-1$  (ijkpunt-1), dan vergroot dit de kans op een baan, 12 maanden later.

**Tabel 4**  
Multivariate modellen voor de kans op werk en de hoogte van het loon 12 maanden na ontslag

	Model 1 Kans op werk (or; logit)	Model 2 Maandloon $t+12$ (log; OLS)	Model 3 Maandloon $t+12$ (log; Heckman)
Geen ontslag	Ref	Ref	Ref
Faillissement	0,09***	-0,06***	-0,07***
UWV bedrijfseconomisch	0,08***	-0,11***	-0,13***
Collectief ontslag	0,12***	0,06**	0,05
45 jaar en ouder	Ref	Ref	Ref
35-44 jaar	1,60***	0,02	0,02
34 jaar en jonger	1,75***	-0,06***	-0,06***
Autochtonen	Ref	Ref	Ref
Westerse allochtoon	0,71**	0,02	0,02
Niet-westerse allochtoon	0,45***	-0,10***	-0,10***
Vrouwen	Ref	Ref	Ref
Mannen	1,31***	0,11***	0,11***
Werkgelegenheidsgroei in sector ( $t-1$ jr)	1,20***	0,01***	0,01***
Werk in zelfde sector		Ref	Ref
Werk in andere sector		0,00	0,00
Werk in andere sector * ontslag		0,00	0,00
Loon in $t-1$ jr (in log)		0,56***	0,57***
$\lambda$			-1,01***
Constante term	22,83***	3,51***	3,49***
<i>Baankans</i>			
Geen ontslag			Ref
Faillissement			-1,12***
UWV bedrijfseconomisch			-1,35***
Collectief ontslag			-1,01***
Vrouwen			Ref
Mannen			0,10*
Leeftijd			0,01
Leeftijd kwadraat			0,00*
Loon in $t-1$ jr (in log)			0,29***
Constante term			-0,31
N	9 375	7 771	8 844,00 <sup>1)</sup>
Pseudo R2	0,19	0,47	nvt

Bron: SSB.

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

<sup>1)</sup> Het aantal waarnemingen is kleiner dan het aantal in model 1, omdat het model voor het loon in  $t-12$  controleert. Deze ontbreekt voor mensen zonder baan op dat moment.

Model 2 (tabel 4) rapporteert de schattingsresultaten van de loonvergelijking. Daarin wordt het loon van werkenden gerelateerd aan de vorm van het bedrijfseconomisch ontslag, waarbij niet-ontslagenen de referentiegroep zijn. In de analyses is ook gecontroleerd voor verandering van sector. Dit wordt gedaan omdat eerdere studies hebben aangegeven dat bij veranderingen van sector een deel van het sectorspecifieke menselijk kapitaal verloren gaat.

De gegevens laten zien dat werknemers die slachtoffer zijn geweest van een faillissement of van UWV-bedrijfseconomisch ontslag 12 maanden na ontslag een significant lager loon verdienen ook al wordt gecontroleerd voor het loon 12 maanden voor ontslag. Het effect van het ontslag bedraagt respectievelijk 6% en 11%. In tegenstelling tot de verwachting wordt in onze analyse geen significant effect van sectorverandering op het loon gevonden.

In model 2 zou er sprake kunnen zijn van selectiviteit. Immers, vooral de meest kansrijke werklozen vanwege bedrijfseconomisch ontslag zullen er in slagen een baan te vinden, hetgeen tot een onderschatting van het effect van ontslagwerkloosheid op het loon zou kunnen leiden. In model 3 (tabel 4) worden de resultaten van een Heckman's selectiemodel gepresenteerd, waarbij de loonvergelijking en de kans op betaalde arbeid simultaan zijn geschat (zie Heckman en Robb, 1985; Gregg en Tominey, 2005). De significante selectieterm ( $\lambda$ ) in de loonvergelijking suggereert dat er inderdaad sprake van een selectie-effect. Het geschatte effect van ontslagwerkloosheid op het loon is als volgt: ontslagenen door faillissement verdienen 7% minder en UWV ontslagenen 13% minder loon. Voor de collectief ontslagenen wordt echter een positief looneffect gevonden. Dit wijst er op dat collectief ontslag een gunstig effect heeft op de allocatie van arbeid.

## 6. *Conclusie*

In dit artikel zijn wij ingegaan op de gevolgen van bedrijfseconomisch ontslag voor de participatie en het loon. Wij sluiten hiermee aan bij recente ontwikkelingen in de economische literatuur, en maken gebruik van de bij het CBS nieuw ontwikkelde datainfrastructuur. Focussen op bedrijfseconomisch ontslag is interessant, omdat, vergeleken met andere vormen van werkloosheid de reden voor ontslag niet gerelateerd is aan de eigen schuld van de ontslagene. De door ons gebruikte databronnen lenen zich dus goed voor het meten van de gevolgen van ontslag op de depreciatie van het menselijk kapitaal: *signaling* van de productiviteit van de werknemer speelt geen rol bij deze vorm van ontslag.

De CBS data maken het mogelijk om drie vormen van bedrijfseconomisch ontslag te onderscheiden: ontslag via UWV, vanwege faillissementen en collectieve uitstroom. Onze analyses hebben laten zien dat ontslagwerklozen een kleinere participatiekans hebben in de jaren na ontslag vergeleken met een controle groep. Bovendien hebben onze analyses laten zien dat ontslag via UWV of als gevolg van

een faillissement significante negatieve gevolgen heeft voor het loon dat iemand na het ontslag verdient. Echter, voor de collectief ontslagenen die weer werk vinden wordt een positief looneffect gevonden. Dit wijst er op dat collectief ontslag een gunstig effect heeft op de allocatie van arbeid.

Het spreekt voor zich dat de hier gepresenteerde analyses slechts een eerste aanzet zijn, en dat wij nog niet alle mogelijkheden van de data hebben benut. In vervolganalyses kan rekening worden gehouden met de duur van de werkloosheid; immers indien er inderdaad sprake is van depreciatie van menselijk kapitaal tijdens een periode van ontslagwerkloosheid, dan kan worden verwacht dat dit effect groter zal zijn naarmate de duur van werkloosheid langer is. Ook zijn onze multivariate analyses nu nog beperkt tot de eerste 12 maanden na werkloosheid. Het zou interessant zijn om te kijken in welke mate de loonverschillen 2 of 3 jaar na ontslag blijven bestaan.

Dat sectorverandering na werkloosheid in onze analyses geen significant effect heeft op het loon is opmerkelijk, omdat eerder studies hebben aangegeven dat het veranderen van sector als gevolg van werkloosheid tot verlies van specifiek menselijk kapitaal leidt. Vervolganalyses zouden hierop dieper in kunnen gaan. Dit zou bijvoorbeeld kunnen door de sectoren te karakteriseren naar de mate waarin het menselijk kapitaal in de sector specifiek of generiek is.

Een mogelijke kritiekpunt op de literatuur over de 'displaced worker' waar deze bijdrage op voortbouwt is dat er ook bij ontslagwerkloosheid sprake zou kunnen zijn van zelfselectie. Daarvan zou sprake zijn als werknemers op een massaontslag anticiperen en de meest kansrijke werknemers al voor de ontslag-golf hun bedrijf verlaten. Het monitoren van stromen van werknemers in en uit bedrijven in de maanden voor massaontslag zou hier meer zicht op kunnen werpen.

## 7. *Discussie*

Onderzoekers kunnen naast enquêtes en langlopende panels in toenemende mate beschikken over administratieve gegevens, in het bijzonder het in dit hoofdstuk gebruikte Sociaal Statistisch Bestand (SSB). Op het gebied van arbeidsmarktonderzoek hebben enquêtes als het OSA-Arbeidsaanbodpanel en de Enquête Beroepsbevolking (EBB) het voordeel dat ze subjectieve informatie over persoonlijke competenties en zoekgedrag naar werk bevatten. Echter, het probleem van de algemeen teruglopende respons en paneluitval, met name in Nederland, wordt groter (Stoop, 2005). Inspanningen vergroten om zoveel mogelijk respondenten over de streep te trekken herbergt het gevaar dat de niet-responderende groep steeds selectiever wordt. Daarnaast hebben enquêtes mogelijk te maken met geheugenartefacten doordat respondenten gebeurtenissen niet precies in de tijd plaatsen, al blijkt dit in de praktijk mee te vallen (Manzoni, 2010). Registraties hebben deze beperkingen niet. We zien in het SSB bijvoorbeeld exact wanneer banen ophouden te bestaan en welke betalingen

worden gedaan aan welke werknemers. UWV registraties leveren precieze informatie over wie, waar, wanneer wordt ontslagen. Registraties zijn echter meestal weer niet opgezet voor sociaalwetenschappelijk onderzoek. In het geval van baanbeëindigingen zijn we afhankelijk van de werkgever die bijvoorbeeld bij een faillissement weinig belang heeft de datum van de beëindiging door te geven.

Net als de 'enquêtewerkelijkheid' dwingt de 'administratieve werkelijkheid' onderzoekers eveneens waakzaam te blijven en deze niet op voorhand op te vatten als dé objectieve beschrijving van de werkelijkheid. We hebben gezien dat het onderzoek er vaak niet eenvoudiger op wordt, ook al bieden deze data vele nieuwe mogelijkheden door het integrale en longitudinale karakter en doordat bronnen over allerlei thema's zijn te koppelen. Het meer combineren en integreren van enquêtes en administratieve gegevens zal toekomstig sociaal-wetenschappelijk onderzoek dichter bij de 'werkelijke werkelijkheid' brengen.

## Referenties

Abbring, J., van den Berg, G., Gautier, P., van Ours, J., van Lomwel, G. & Ruhm, C. (2002). Worker Displacement in the US and the Netherlands, in: P. Kuhn (ed.), *Losing work, moving on: international perspectives on worker displacement*, W.E. Upjohn Institute for Employment Research, Kalamazoo, Michigan, 105–194.

Albrecht, J., Edin, P., Sundstrom, M., Vroman, S. (1999), Career interruptions and subsequent earnings: A reexamination using Swedish data, *The Journal of Human Resources*, 34 (2), 294–311.

Arts, C. H., Bakker, B. F. M., & Van Lith, E. (2000). Matching administrative registers and household surveys. *Netherlands Official Statistics*, 15 (Summer; Special Issue: Integrating administrative registers and household surveys), 16–22.

Arts, C. H., & Hoogteijling, E. M. J. (2002). Sociaal Statistisch Bestand 1998 en 1999. *Sociaal-Economische Maandstatistiek*, 12, 13–21.

Arulampalam, W. (2001), Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages, *The Economic Journal*, 111(475), F585–F606.

Bakker, B. F. M., Tijdens, K. G., & Winkels, J. W. (1999). Explaining gender wage differences. *Netherlands Official Statistics*, 14(Winter), 36–41.

Bakker, B. F. M. (2008). De stand van het Sociaal Statistisch Bestand, *Bevolkingstrends* 56(2), 14–18

Bakker, B. F. M. (2009). *Trek alle registers open!* Rede uitgesproken van het ambt van bijzonder hoogleraar Methodologie van registerdata voor sociaal-wetenschappelijk onderzoek: Vrije Universiteit Amsterdam.

Bloemendal, C. (2010). Wie verliezen hun baan bij faillissementen? *Sociaaleconomische trends* (2e Kwartaal), 49–53.

Burda, M. & A. Mertens (2001), Estimating wage losses of displaced workers in Germany, *Labour Economics*, 8, 15–41.

Corpeleijn, A. (2009). Werkhervatting na ontslag: een vergelijking van oudere en jongere werknemers. *Sociaaleconomische Trends*(2), 35–40.

Couch, K. (2001), Earnings Losses and Unemployment of Displaced Workers in Germany, *Industrial and Labor Relations Review*, 54(3), 559–572.

Edin, P., Gustavson, M. (2008), Time out of work and skill depreciation, *Industrial & Labor Relations Review*, 61 (2), 163–180.

Gangl, M. (2006) Scar Effects of Unemployment: An Assessment of Institutional Complementarities, *American Sociological Review*, 71(6), 986–1013.

Gibbons, R., Katz, L. (1991), Layoffs and lemons, *Journal of Labor Economics*, 9(4), pp.351–380.

Gregg, P. & E. Tominey (2005), The Wage Scar From Male Youth Unemployment, *Labour Economics*, 12(4), 487–509.

Gregory, M. & R. Jukes (2001), Unemployment and Subsequent Earnings: Estimating Scarring among British Men 1984–94, *The Economic Journal*, 111(475), F607–F625.

Hamermesh, D. (1987), The Costs of Worker Displacement, *The Quarterly Journal of Economics*, 102(1), 51–76.

Heckman, J. & R. Robb (1985), Alternative Methods for Evaluating the Impact of Interventions: An Overview, *Journal of Econometrics*, 30(1–2), 239–267.

Hijzen, A., R. Upward & P. Wright (2010), The Income Losses of Displaced Workers, *The Journal of Human Resources*, 45(1): 243–269.

Jacobson, L., LaLonde, R., Sullivan, D. (1993), Earnings Losses of Displaced Workers, *The American Economic Review*, 83(4), 685–709.

- Kletzer, L. (1998), Job Displacement, *Journal of Economic Perspectives*, 12, 115–136.
- Koeber, C., Wright, D. (2006), Gender differences in the reemployment status of displaced workers human capital as signals that mitigate effects of bias, *Journal of Socio-Economics*, 35 (5), 780–796
- Kriechel B (2010), Displaced Workers, Unemployed and Vocational Education and Training. In: Penelope Peterson, Eva Baker, Barry McGaw, (Editors), *International Encyclopedia of Education*. volume 8, pp. 275–279. Oxford: Elsevier.
- Kriechel, B. & G. Pfann (2005). The role of specific and general human capital after displacement, *Education Economics*, 13(2), 223–236,
- Kriechel, B. (2003), *Heterogeneity among displaced workers*, Ph.D. thesis, Maastricht University.
- Kuijvenhoven, L., & Scholtus, S. (2010). *Estimating accuracy for statistics based on register and surveys*. Statistics Netherlands, Den Haag: <http://www.cbs.nl/NR/rdonlyres/5407AA5A-A198-4E2B-857D-87169E5408C3/0/201007x10pub.pdf>.
- Manzoni, A. (2010) *Labor mobility patterns over the life course. A comparison of retrospective and prospective data in different labor markets*, PhD-thesis, Tilburg University.
- Mooi-Reçi. (2008). *Unemployed and scared for life? Longitudinal analyses of how unemployment and policy changes affect re-employment careers and wages in the Netherlands, 1980–2000*. Amsterdam: Vrije Universiteit.
- Neal, D. (1995), Industry-Specific Human Capital: Evidence from Displaced Workers, *Journal of Labor Economics*, 13(4), 653–677.
- Pissarides, C. (1992), Loss of skill during unemployment and the persistence of employment shocks, *The Quarterly Journal of Economics*, 107(4), 1371–1391.
- ROA (2009). *De arbeidsmarkt naar opleiding en beroep tot 2014*. ROA-R-2009/5, Universiteit Maastricht.
- Scheele, D., Theeuwes, J. J. M., & de Vries, G. J. M. (2007). *Arbeidsflexibiliteit en ontslagrecht*. Amsterdam: Amsterdam University Press.
- Scheele, D., van Gaalen, R., & van Rooijen, J. (2008). *Werk en inkomsten na massaontslag: De zekerheid is niet van de baan*. Amsterdam: Amsterdam University Press.



Song, Y. (2009), Training, Technological Changes, and Displacement, *Journal of Labor Research*, 30, 201–218.

Schils, T., Fouarge, D., Kerkhofs, M. (2006), *Loon en werk na werkloosheid*, Tilburg: OSA publicatie A221.

Song, Y. (2009), Training, Technological Changes, and Displacement, *Journal of Labor Research*, 30, 201–218.

Stoop, I. A. L. (2005). *The hunt for the last respondent. Non-response in sample surveys*. The Hague: Social and Cultural Planning Office of the Netherlands.

Van Gaalen, R., & Van Rooijen, J. (2009). CBS Berichten: Participatie en inkomsten na collectief en individueel ontslag om bedrijfseconomische redenen. *Tijdschrift voor Arbeidsvraagstukken*, 25(2), 201–206.

## Noten

- 
- <sup>1)</sup> Dit onderzoek is tot stand gekomen in het kader van de CBS/ROA samenwerkingsproject “Scholing en mobiliteit”.
  - <sup>2)</sup> Dit zou het gevolg kunnen zijn van het feit dat massaontslag ook werknemers omvat die vrijwillig het bedrijf verlaten.
  - <sup>3)</sup> Het verlies aan loon ligt bij het vrijwillig terugtrekken van de arbeidsmarkt wel hoger bij vrouwen.
  - <sup>4)</sup> Dit zou er op kunnen wijzen dat er relatief veel bedrijven failliet gaan die recentelijk zijn gestart.
  - <sup>5)</sup> In vervolganalyses overwegen wij om de ontslagenen te ‘matchen’ aan niet-ontslagenen met vergelijkbare kenmerken.
  - <sup>6)</sup> Een odds-ratio onder (boven) de waarde 1 betekent dat het kenmerk wordt geassocieerd met een kleinere (grotere) baankans.



# *Opleidingsniveau in registers: een toets van de validiteit via loonfuncties*

Aslan Zorlu (UvA)

## *1. Inleiding*

Een groot aantal studies laat zien dat opleidingsniveau een grote voorspellende kracht heeft voor de positie van mensen in de samenleving. Zo weten we dat hoger opgeleide mensen vaak gezonder zijn, banen van een hoger niveau hebben en meer verdienen. Gedrag patronen en attitudes van hoger opgeleide personen ten aanzien van investeringen, demografische verschijnselen en politiek verschillen sterk van die van laagopgeleiden. Voor onderzoekers is opleidingsniveau dan ook een onmisbare variabele. Een zwakke schakel bij de registers is de informatie over opleidingsniveau. Immers, anders dan voor belangrijke demografische en arbeidsmarktkenmerken van individuen, bestaat er geen goed register waarin het opleidingsniveau van de Nederlandse bevolking is opgenomen.

De bekostigde opleidingen van studenten en leerlingen worden weliswaar door afzonderlijke onderwijsinstellingen geregistreerd en geleverd aan de Informatie Beheergroep, maar dat is iets van de laatste twintig tot dertig jaar. Er is geen registerinformatie over opleidingen die in het verdere verleden gevolgd zijn. Dit betekent dat anno 2010 het opleidingsniveau voor mensen boven de dertig à veertig minder goed is vast te stellen uit registers. Het probleem van een ontbrekende opleidingvariabele zal overigens op termijn verdwijnen naarmate nieuwe lichtingen aan onderwijs deelnemen en de populatie van huidige ouderen zal afnemen. Dat zal echter nog tientallen jaren duren. Een ander probleem is dat in de onderwijsregisters alle in het buitenland gevolgde opleidingen ontbreken. Deze lacune heeft in het bijzonder consequenties voor onderzoek naar immigranten omdat kennis over buitenlandse opleidingen onontbeerlijk is om de mate van *skill transferability* te kunnen meten (Friedberg 2000). Bovendien internationaliseert het hoger onderwijs in toenemende mate waardoor de informatie over in het buitenland gevolgde opleidingen steeds belangrijker wordt.

Het CBS heeft recent stappen gezet om een opleidingsvariabele te produceren door middel van het combineren van een aantal registers en surveys, zoals de jaarlijkse Enquête beroepsbevolking (EBB), het Centraal Register Inschrijvingen in het Hoger Onderwijs (CRIHO), het Examen Resultaten Register (ERR), het studiefinancieringregister (WSF) en Onderwijsnummerbestanden voor het Voortgezet Onderwijs en Middelbaar Beroepsonderwijs (Bakker et al., 2008). Dit resulteerde in een

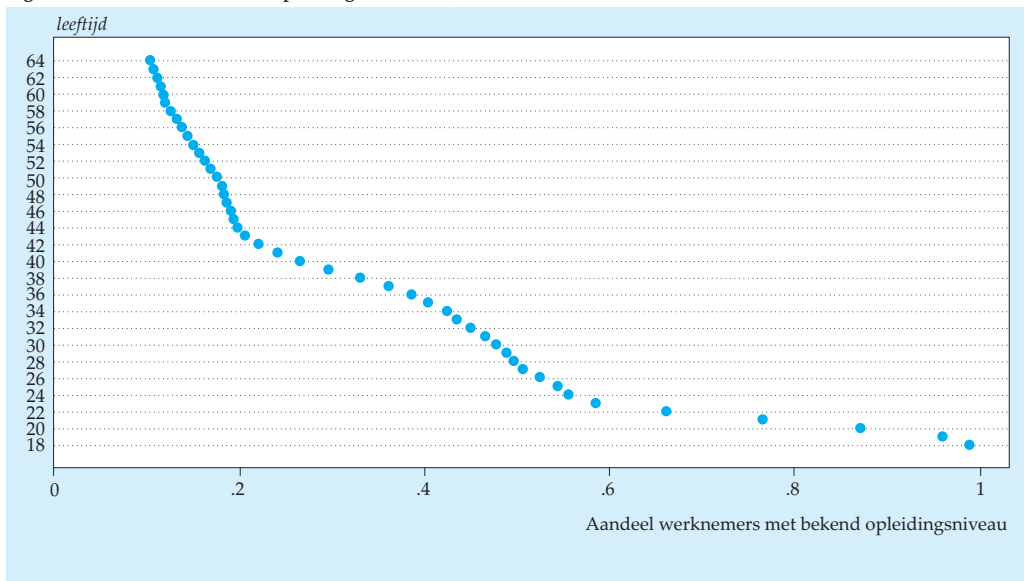
opleidingsvariabele, opleidingsniveau genoemd, die deel uitmaakt van het Sociaal Statistisch Bestand (SSB) en in 2008 45% van de bevolking dekt. Vanzelfsprekend is deze variabele selectief en dat levert methodologische complicaties op om het opleidingsniveau vooral voor oudere mensen goed te kunnen vaststellen. Dat is nog gecompliceerder voor niet-westerse immigranten die vaker de Nederlandse taal onmachtig zijn en daardoor minder vaak aan algemene surveys deelnemen. Daarnaast hebben ze een groot deel van hun opleiding in hun herkomstland gekregen wat per definitie buiten het domein van de Nederlandse registers valt.

In dit artikel toetsen we de validiteit van de opleidingsvariabele uit het SSB, in het bijzonder voor immigranten. Daarbij gebruiken we ook de integrale gegevens over banen en demografische gegevens over personen uit het SSB.

## 2. *Structuur van het opleidingsregister*

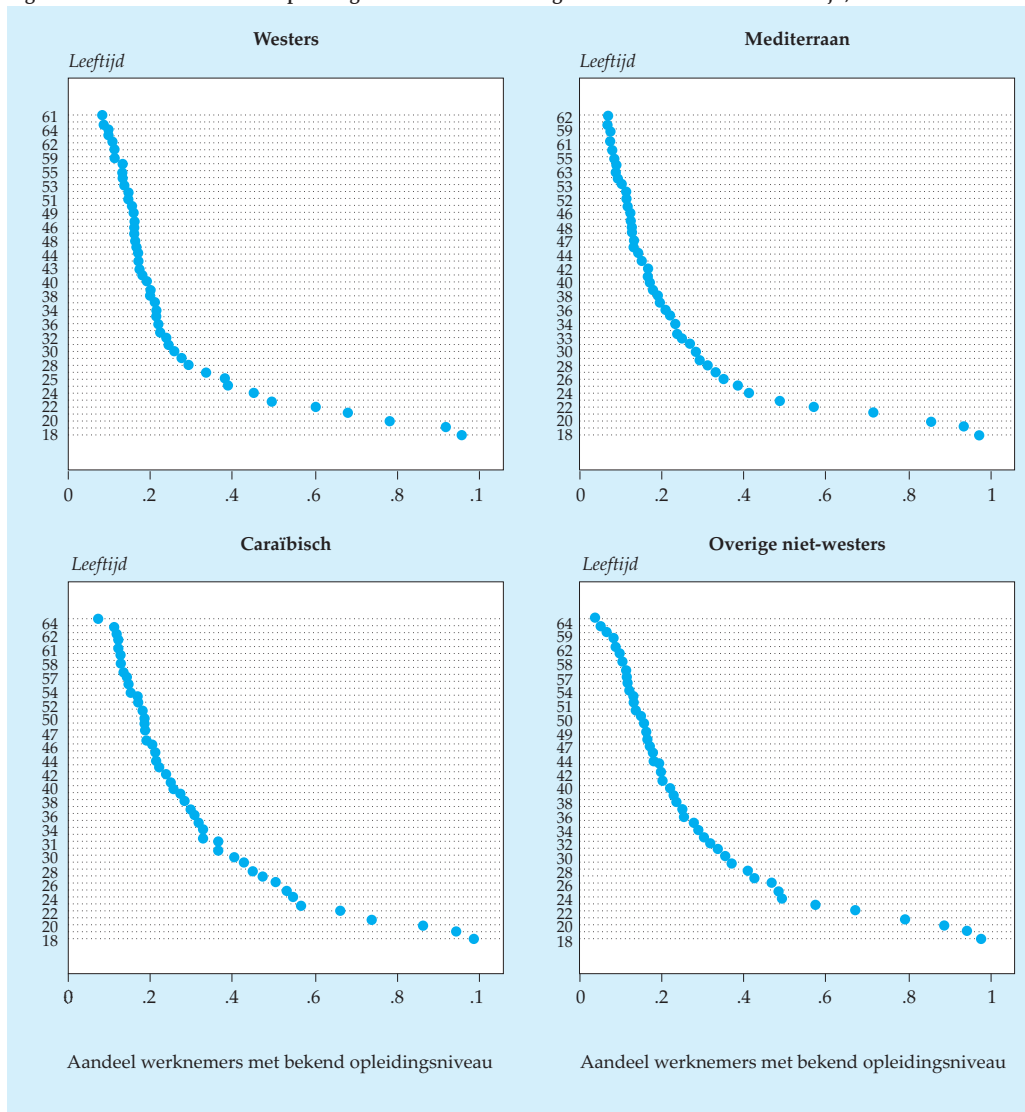
Een analyse van de opleidingsvariabele uit het SSB laat zien dat het percentage van werknemers voor wie opleidingsinformatie beschikbaar is, sterk varieert naar leeftijd. Het opleidingsniveau van jongeren is vaker bekend dan dat van oudere werknemers. Voor 18- en 19-jarige werknemers is het opleidingsniveau nagenoeg voor iedereen bekend. Daarna daalt het percentage van werknemers met een bekend opleidingsniveau naarmate de leeftijd hoger is. Voor iets minder dan 60 procent van de 23-jarigen is het opleidingsniveau bekend. Onder werknemers ouder dan 40 jaar, is het aandeel met een gemeten opleidingsniveau lager dan 20 procent.

**Figuur 1. Beschikbaarheid van opleidingsinformatie, 2006**



De scheve leeftijdsverdeling van bekende opleidingsniveaus is gerelateerd aan de registraties en enquêtes die zijn gebruikt om het opleidingsregister te creëren. Daarbij is de rol van het CRIHO en de recentelijk ingevoerde Onderwijsnummerregistraties belangrijk. Vooral dank zij de laatstgenoemde registraties is voor een hoog percentage van werknemers onder de 20 jaar een opleidingsniveau bekend. Het CRIHO garandeert dat voor alle werknemers onder de 40 tot 45 jaar met een in Nederland genoten HBO- of WO-opleiding er voldoende opleidingsinformatie is. De bijdrage van de Enquête beroepsbevolking (EBB, 1996–2009) is in principe

Figuur 2. Beschikbaarheid van opleidinginformatie voor eerste generatie allochtonen naar leeftijd, 2006

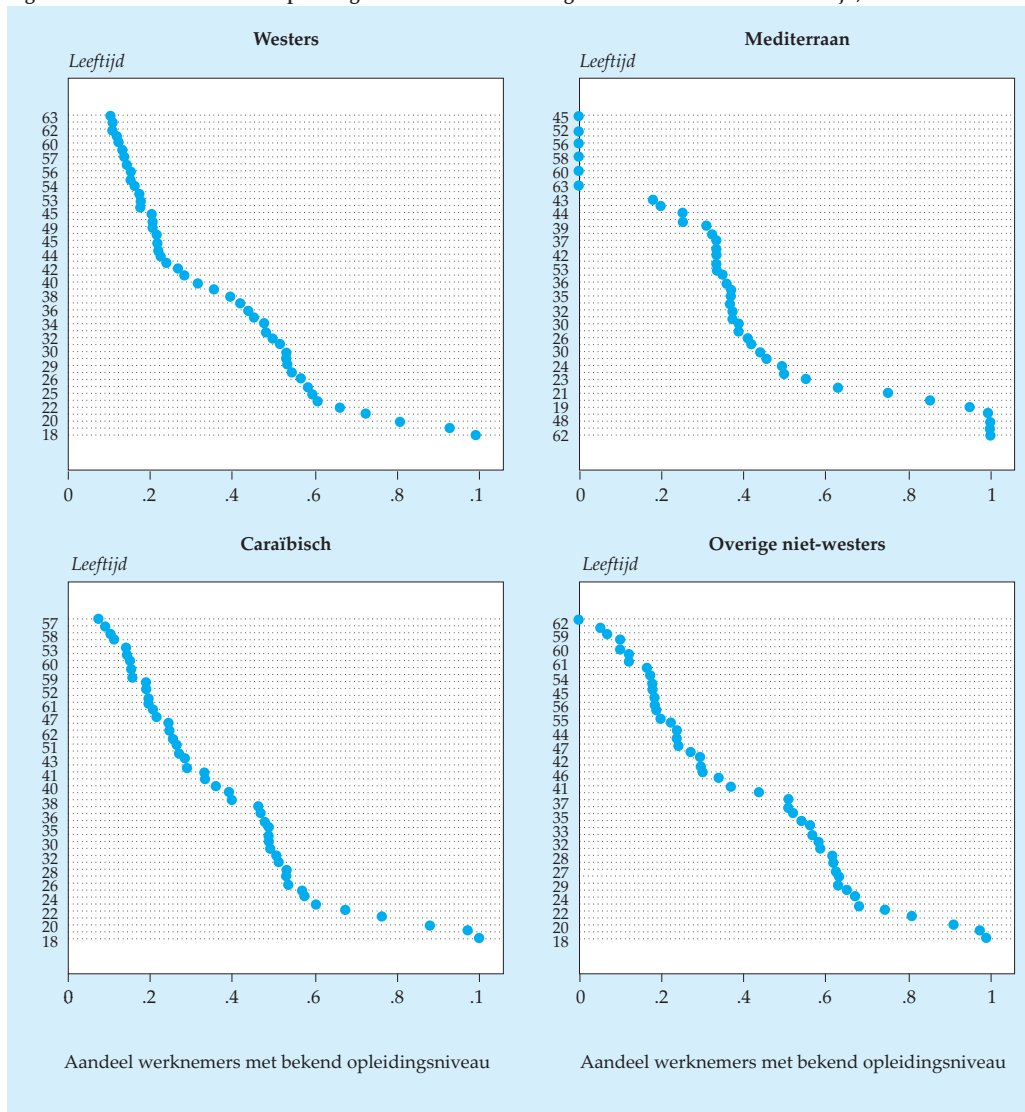


goed verdeeld over de leeftijdscategorieën aangezien de EBB een na herweging representatieve steekproef is uit de bevolking van 15 jaar en ouder. Omdat een deel van de EBB oudere jaargangen betreft dekt deze bron nog beter de oudere leeftijdscategorieën af.

De vraag is of de dekkingsgraad naar leeftijd van de SSB-variabele opleiding voor de gehele werknemerspopulatie afwijkt van die van de werkzame immigranten (eerste generatie) en hun werkzame kinderen (tweede generatie). Een vergelijking onder werknemers van niet-westerse immigranten met autochtonen<sup>1</sup> en tweede generatie allochtonen laat een afwijkend patroon zien. Voor allochtone werknemers van de eerste generatie is het opleidingsniveau aanzienlijk minder vaak bekend dan voor de niet-immigranten (Figuur 2). Dat suggereert dat immigranten in de EBB- steekproeven niet evenredig zijn vertegenwoordigd. Deze uitkomst is echter niet onverwacht gezien een relatief lage respons van immigranten in enquêtes. Het is immers bekend dat de taalbarrière onder de eerste generatie laaggeschoolde immigranten een belangrijke belemmering is en blijft om aan enquêtes deel te nemen. Daarnaast heeft een groot deel van de immigranten waarschijnlijk hun opleiding in het herkomstland gekregen waardoor ze in de opleidingsregisters niet voorkomen. Voor de tweede generatie is de opleiding bijna even vaak bekend als voor de hele werknemerspopulatie omdat ze vaak onder de leeftijd 35 zijn geconcentreerd en waarschijnlijk hun opleiding in Nederland hebben gevolgd (Figuur 3).

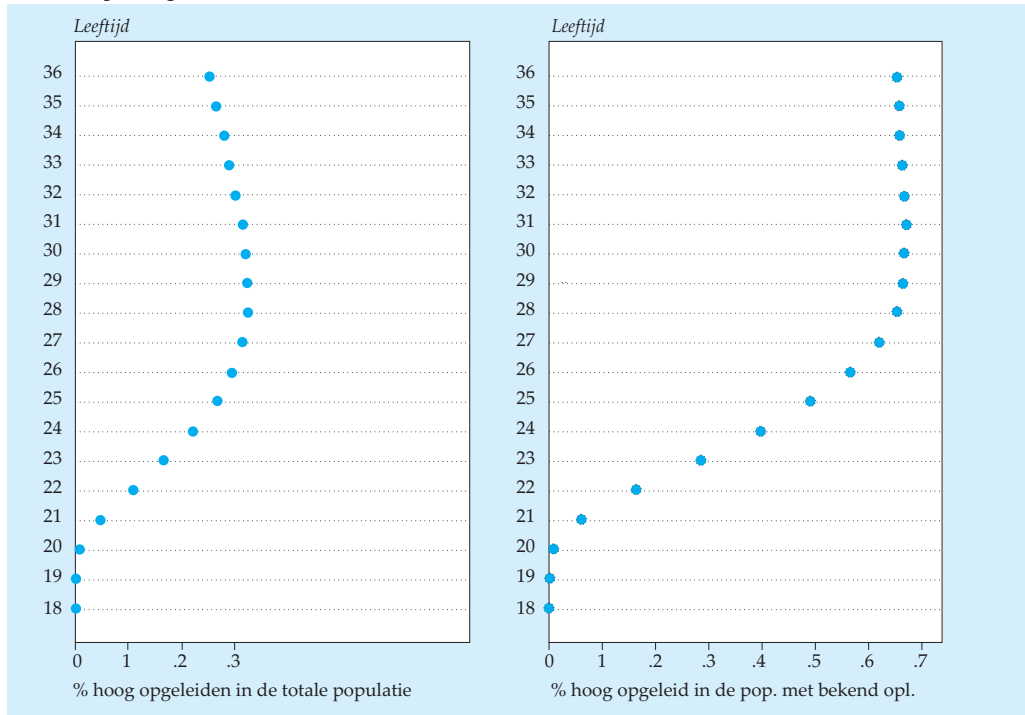
Om de validiteit van het opleidingsniveau uit het SSB te onderzoeken en de consequenties van het ontbreken van opleidingsinformatie voor (specifieke) migrantengroepen in kaart te brengen concentreren we ons op werknemers die tussen 1970 en 1988 zijn geboren (18–36 jaar in 2006). Het opleidingsniveau uit het SSB bevat in 2006 vooral informatie uit het CRIHO, de ERR en de EBB. De Onderwijsnummersbestanden zijn dan nog van te recente datum om een substantiële bijdrage aan het register te leveren voor de populatie werknemers. Dat betekent wel dat we over opleidingsinformatie beschikken voor alle werknemers die een diploma hebben voor een HBO- of WO- studie althans als ze hun opleiding in Nederland hebben gevolgd. De selectie van deze geboortecohorten biedt een geschikt raamwerk waarbinnen het effect van een hoger onderwijsdiploma op het loon kan worden bepaald. Daarnaast kunnen mogelijk meer zuivere loonverschillen tussen herkomstgroepen worden bepaald na een adequate correctie voor opleidingsverschillen tussen de groepen. Het is immers bekend dat niet-westerse allochtonen vaker laag opgeleid zijn dan autochtonen en westerse migranten. Voor niet-hoog opgeleide werknemers kan de opleidingsinformatie uiteraard onvolledig zijn. Daarom komt de nadruk te liggen op de relatieve positie van hoogopgeleiden ten opzichte van niet-hoogopgeleiden. Voor deze groep van werknemers kijken we eerst naar de kans dat een werknemer een hoog onderwijsdiploma heeft. Dat doen we op twee manieren: eerst kijken we naar het aandeel van werknemers met een onder-

Figuur 3. Beschikbaarheid van opleidinginformatie voor tweede generatie allochtonen naar leeftijd, 2006



wijsdiploma in de totale populatie van werknemers tussen 18 en 36 jaar (Figuur 4). Vervolgens kijken we naar het aandeel van werknemers met een onderwijsdiploma in de populatie van werknemers waarvoor opleidinginformatie beschikbaar is. Figuur 4 laat zien dat de kans op een diploma vrijwel nihil is voor jongeren onder 21. Deze kans loopt sterk op tot 28 jaar, de leeftijd waarbij de top bereikt wordt in het aandeel hoog opgeleiden. Theoretisch gezien zouden jongeren hun studie moeten kunnen afmaken rond de leeftijd van 21 en 22, maar een studie duurt in de praktijk kennelijk langer.

Figuur 4. Aandeel werknemers met hoger opleiding in de totale populatie en in de populatie met bekend opleidingsniveau, 2006



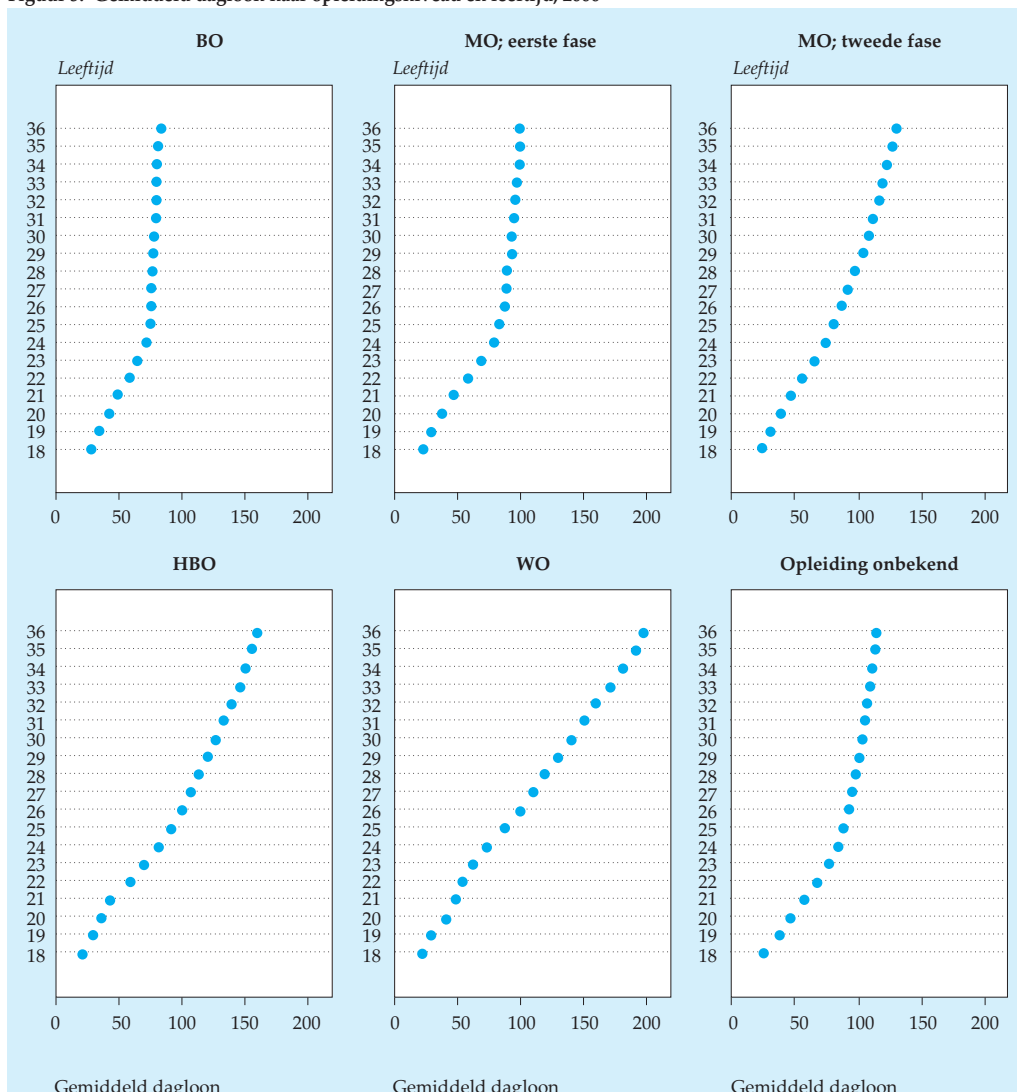
Het aandeel van hoogopgeleiden in de totale populatie bereikt zijn maximum op een niveau boven de 30 procent voor werknemers die 28 en 29 jaar zijn (linkerfiguur). Dat neemt daarna af tot 24 procent voor 36 jarigen. Dat zou betekenen dat het aandeel van hoogopgeleiden in de tijd aanzienlijk is gestegen. De rechterfiguur laat zien dat het aandeel van hoogopgeleiden onder werknemers met een bekende opleiding een vergelijkbaar patroon volgt langs de leeftijdslijn hoewel de daling na 29 jaar minder sterk is. Belangrijk is de hoogte van het aandeel hoogopgeleiden in deze populatie. Wanneer de opleiding bekend is, heeft bijna 70 procent van deze werknemers een hoger onderwijsdiploma. Dit hoge percentage maakt nogmaals duidelijk dat het opleidingsregister vrij selectief is ten gunste van hoogopgeleiden hetgeen voor een belangrijk deel is toe te schrijven aan het grote aandeel van de CRIHO in de gebruikte bronnen.

### 3. Lonen

Allereerst wordt nagegaan hoe het loonniveau van werknemers tussen 18 en 36 met verschillende opleidingsniveaus varieert naar leeftijd. Figuur 5 laat de ontwikkeling van lonen zien naar leeftijd en opleidingsniveau. Zoals te verwachten stijgt het loonniveau sterk voor werknemers met basis- of middelbaar onderwijs, 1<sup>e</sup>

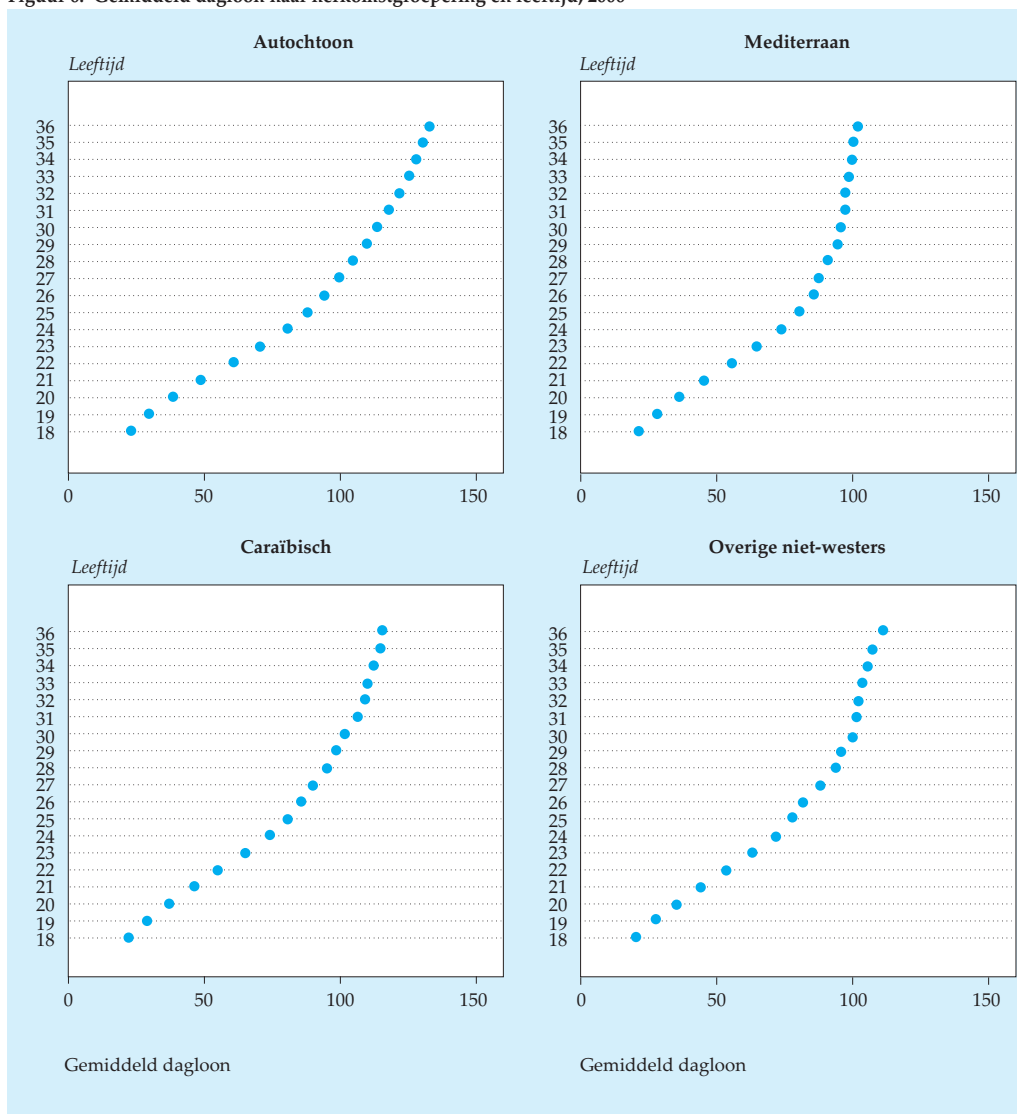


Figuur 5. Gemiddeld dagloon naar opleidingsniveau en leeftijd, 2006



fase (bijv. VMBO, onderbouw HAVO en VWO) tot de leeftijd van 24 jaar. Daarna stagneert de loonontwikkeling voor deze groep. Voor werknemers met een diploma in het middelbaar onderwijs, 2<sup>e</sup> fase (diploma VWO, HAVO, volledig MBO) gaat de loonontwikkeling met iets afnemende snelheid verder. Voor afgestudeerden van het HBO vindt er een kleine sprong plaats in de loonontwikkeling rond de afstudeerleeftijd van 21–22 jaar. Daarna stijgen de lonen in een rustiger tempo bij oplopende leeftijd. Werknemers met een universitair diploma hebben het sterkst stijgende loonprofiel. Het loonprofiel van werknemers zonder opleidingsinformatie lijkt het meest op die van werknemers met een middelbaar onderwijsdiploma, 1<sup>e</sup> fase

Figuur 6. Gemiddeld dagloon naar herkomstgroepering en leeftijd, 2006



Figuur 6 laat de loonprofielen zien per herkomstgroep. Een interessant verschijnsel is dat alle herkomstgroepen een vergelijkbare startpositie hebben. Rondom de leeftijd van 25 jaar tekenen zich grote verschillen af. Vanaf deze leeftijd stagneert de loonontwikkeling van de bevolking van Mediterraanse herkomst en overige niet-westerse allochtonen, en wat in mindere mate voor werknemers uit de Caraïben. Voor autochtone werknemers is de loonontwikkeling duidelijk sterker. Natuurlijk komt het merendeel van deze verschillen voort uit de verschillen in opleidingsniveaus en andere achtergrondkenmerken. Daarom zetten we onze analyse voort

met regressie-analyses waarbij we daglonen corrigeren voor waarneembare verschillen in relevante kenmerken van werknemers.

#### 4. *Een regressieanalyse met de SSB-opleidingsvariabele*

In paragraaf 2 is duidelijk geworden dat de kans op een bekende opleiding in het SSB nauw samenhangt met leeftijd en herkomst. Zoals eerder gezegd berust de SSB-opleidingsvariabele op informatie uit registers en surveys (steekproeven). Er wordt daarbij zoveel als mogelijk gebruik gemaakt van de registerbronnen, en daar waar deze tekortschieten worden de surveys te hulp geroepen. De selectiviteit naar o.a. leeftijd en herkomst wordt gecorrigeerd door aan de records uit het steekproefdeel een gewicht mee te geven. Dit gewicht is zo ontworpen dat de combinatie van register- en surveyrecords voor verschillende achtergrondkenmerken (bv. leeftijd, geslacht en herkomst) een representatieve afspiegeling geven van de gehele populatie. De verzameling van mensen met een bekend opleidingsniveau kan dus worden opgehoogd tot een niveau van de gehele populatie. Het spreekt vanzelf dat toepassen van dit gewicht voor kleinere subpopulaties niet automatisch leidt tot een geldige uitkomst. Bovendien is het werkelijke opleidingsniveau van de gehele populatie en de subpopulaties waaruit deze is samengesteld niet bekend.

Om de consequenties van het ontbreken van opleidingsniveaus in de registers te onderzoeken gebruiken we de populatie van werknemers tussen 18 en 36 jaar. Voor deze populatie schatten we standaard loonfuncties waarbij (de logaritme van) het dagloon wordt voorspeld door demografische en sociaal-economische kenmerken van personen. Deze kenmerken omvatten naast opleiding de volgende voorspellende variabelen: geslacht, leeftijd (ook het kwadraat en de derde macht), de zes grote herkomstgroepen onderscheiden naar generatie, anciënniteit (ook het kwadraat), al dan niet in deeltijd werken en deeltijdfactor, al dan niet flexibel werken, type huishouden, en 10 bedrijfssectoren. Daarvoor volgen we de volgende strategie: eerst schatten we een model waarin de opleidingsvariabele niet opgenomen is (model I); ten tweede nemen we drie dummyvariabelen op voor middelbare, HBO- en WO- opleiding, en een controlevariabele voor wie het opleidingsniveau onbekend is (model II); ten derde schatten we de loonfunctie voor de populatie van werknemers voor wie het opleidingsniveau bekend is (model III); als laatste schatten we model III voor dezelfde populatie, maar dan met het gewicht dat is ontworpen om deze populatie op te hogen tot de totale populatie (model IV).

De geschatte modellen zijn weergegeven in tabel 1. Aangezien we geïnteresseerd zijn in het effect van opleidingsniveau en variatie in lonen tussen herkomstgroepen als gevolg van het al dan niet corrigeren voor opleidingsniveauverschillen, presenteren we alleen de coëfficiënten voor dummyvariabelen voor de opleidingsniveaus, eerste en tweede generatie herkomstgroepen, anciënniteit, geslacht en (niet-lineaire) leeftijd.

**Tabel 1**  
**Standaard loonfuncties geschat met behulp van verschillende modellen, 2006**

	Model I	Model II	Model III	Model IV
<i>Opleidingsniveau (alleen bo = ref.)</i>				
Middelbare opleiding		0,073***	0,068***	0,076***
HBO opleiding		0,244***	0,225***	0,238***
WO opleiding		0,358***	0,337***	0,353***
Opleidingsniveau onbekend		0,059***		
<i>Leeftijd en anciënniteit</i>				
Leeftijd	1,180***	1,102***	1,011***	1,108***
Leeftijd <sup>2</sup>	-0,037***	-0,035***	-0,032***	-0,035***
Leeftijd <sup>3</sup>	0,000***	0,000***	0,000***	0,000***
Anciënniteit	0,047***	0,051***	0,064***	0,050***
Anciënniteit <sup>2</sup>	-0,003***	-0,003***	-0,004***	-0,002***
<i>Geslacht (man = referentiecategorie)</i>				
Vrouw (man = referentiecategorie)	-0,068***	-0,075***	-0,032***	-0,071***
<i>Herkomstgroepering (autochtoon = ref.)</i>				
Marokkaan - Eerste generatie	-0,131***	-0,088***	-0,066***	-0,093***
Marokkaan - Tweede generatie	-0,023***	0,002	0,006	0,052***
Turks - Eerste generatie	-0,138***	-0,094***	-0,072***	-0,092***
Turks - Tweede generatie	-0,036***	-0,009***	-0,015***	-0,016
Surinamer - Eerste generatie	-0,105***	-0,059***	-0,038***	-0,058***
Surinamer - Tweede generatie	-0,034***	-0,007**	-0,002	0,004
Antilliaan - Eerste generatie	-0,124***	-0,081***	-0,081***	-0,061***
Antilliaan - Tweede generatie	-0,011*	-0,002	-0,005	0,002
Overige niet-west - Eerste generatie	-0,111***	-0,068***	-0,089***	-0,085***
Overige niet-west - Tweede generatie	-0,022***	-0,021***	-0,027***	-0,010
Westers - Eerste generatie	-0,002	0,043***	-0,056***	-0,043***
Westers - Tweede generatie	-0,011***	-0,005***	-0,006***	-0,009
N	2 555 663	2 555 663	1 387 385	1 387 385
R <sup>2</sup>	0,637	0,655	0,714	0,663

\* p<0.05; \*\* p<0.01; \*\*\* p<0.001.

Model I: zonder opleiding dummy's; de totale populatie van werknemers.

Model II: met opleiding dummy's; een dummy voor onbekende opleiding.

Model III: met opleiding dummy's; de subpopulatie van werknemers met opleidingsinformatie.

Model IV: met opleiding dummy's; de subpopulatie van werknemers met opleidingsinformatie met gewicht.

Alle modellen omvatten ook variabelen voor al dan niet parttime werken, al dan niet flexibel werken, type huishouden, 10 bedrijfssectoren.

Model 1 geeft de bovengrens van loonverschillen als de parameterschattingen niet worden gecorrigeerd voor opleidingsniveauverschillen. De eerste generatie niet-westers immigranten hebben 11 tot 13 procent lagere lonen dan autochtonen. De loonverschillen tussen autochtonen en tweede generatie is relatief klein: circa 3 procent. Model II corrigeert voor opleidingsverschillen voor het deel van de werknemers dat een bekend opleidingsniveau heeft. Daarnaast is een dummyvariabele opgenomen voor opleidingsniveau onbekend. Dit reduceert de loonachterstand van niet-westerse immigranten tot 7 à 9 procent. De achterstand voor de tweede generatie verdwijnt, is minder groot of wordt statistisch insignificant wanneer er rekening wordt gehouden met opleidingsniveauverschillen. In dit model verdienen werknemers met een HBO- en WO-diploma respectievelijk 24 en bijna 36 procent meer dan laag opgeleide werknemers. De loonwinst is ongeveer 7 procent voor middelbaar opgeleide werknemers. Wanneer we alleen de populatie van werknemers met een bekend opleidingsniveau gebruiken om de loonfunctie te

schatten met model III, zien we dat de loonachterstand kleiner is voor immigranten uit Marokko, Suriname en Turkije dan in model II, terwijl dat gelijk is voor Antilliaanse en hoger is voor overige niet-westerse immigranten. De geschatte coëfficiënten voor de opleidingscategorïeën zijn iets lager dan de coëfficiënten in model II. Model IV wordt toegepast op dezelfde subpopulatie als model III om een loonfunctie te schatten, maar gebruikt daarbij een gewicht. Als het gewicht de eigenschappen van subpopulaties volledig kan verdisconteren, zouden we mogen verwachten dat loonverschillen tussen de herkomstgroepen in model II en model IV van gelijke omvang zijn. Dit blijkt het geval, al nemen we kleine verschillen in coëfficiënten waar voor Antilliaanse, overig niet-westerse en Marokkaanse werknemers. De verschillen zijn echter bijna te verwaarlozen voor Turkse, Surinaamse en westerse immigranten.

Op grond van deze experimenten valt te concluderen dat loonverschillen tussen autochtonen en niet-westerse allochtone groepen aanzienlijk groter zijn wanneer we opleidingsverschillen negeren. Het opleidingsniveau zoals dat in het SSB is opgenomen, helpt de bias in de schattingen behoorlijk te reduceren. De geschatte coëfficiënten voor de opleidingvariabele zijn ongeveer gelijk van omvang in twee modellen waarin in een model een controle variabele voor ontbrekende waarden is opgenomen en in een alternatief model ontbrekende waarden zijn uitgesloten en een gewicht is gebruikt (modellen II en IV). Belangrijk is op te merken dat de variabelen in deze modellen ook een groot deel van de selectiviteit in het SSB-opleidingsregister zouden kunnen elimineren. Dat betekent dat een opleidingvariabele die selectief is, kan worden gebruikt samen met een controlevariabele voor ontbrekende waarden in een model met standaard demografische achtergrondvariabelen. Onze analyse levert verder een indicatie op voor mogelijke consequenties van het uitsluiten van de subpopulatie zonder opleidinginformatie. In dit geval schieten andere achtergrondvariabelen tekort om de selectiviteit in opleidingsregister te elimineren.

## 5. *Conclusie*

Het is evident dat de variabele opleidingsniveau een centrale rol speelt in sociaal-wetenschappelijk onderzoek. Deze cruciale variabele is echter niet gemakkelijk te verkrijgen uit registers. De enorme variatie aan opleidingen en coördinatieproblemen tussen opleidingsinstellingen maken het niet gemakkelijk om opleidingsniveau van iedereen tijdig te registreren en bij te houden. Het CBS heeft recentelijk een opleidingsvariabele samengesteld op basis van beschikbare onderwijsregistraties (zoals het CRIHO en de Onderwijsnummerbestanden) en een reeks surveys (de EBB). De variabele is opgenomen in het Sociaal Statistisch Bestand (SSB). De variabele dekt in 2006 minder dan 30 procent van de potentiële beroepsbevolking. Het dekkingspercentage varieert sterk naar leeftijd: voor jongeren is de opleidinginformatie vaker beschikbaar dan voor ouderen.

In dit artikel werd de validiteit onderzocht van de variabele opleidingsniveau uit het SSB, daarbij ook gebruik makend van de informatie over de banen uit het SSB. Aangezien het CRIHO voor het opleidingsniveau in het SSB in 2006 een heel belangrijke bron is en integrale data over het hoger onderwijs vanaf 1985 bevat, is de keuze gemaakt om de analyses te beperken tot de werknemers tussen 18 en 36 jaar. Op deze manier wordt een raamwerk gedefinieerd waarbinnen met grote zekerheid is vast te stellen of een werknemer afgestudeerd is bij een HBO- of WO-instelling. Op deze subpopulatie is geprobeerd de consequenties van het ontbreken van een deel van de opleidingsniveaus en de validiteit van het wel aanwezige deel in het SSB in kaart te brengen. Hiervoor wordt regressieanalyse toegepast om het loonniveau van werknemers te voorspellen vanuit standaard achtergrondvariabelen als geslacht, herkomstgroepering, leeftijd, opleiding, anciënniteit, de huishoudenscompositie en de sector waarin men werkt.

De resultaten laten zien dat het effect van opleidingsniveau op het dagloon goed geschat kan worden wanneer schattingen worden gecorrigeerd voor de waarnemingen zonder opleidinginformatie. In een alternatieve setting kan het effect van opleiding worden geschat gebaseerd op de subpopulatie waarvoor opleidinginformatie beschikbaar is wanneer het door het CBS ontwikkelde bijbehorende gewicht wordt gebruikt. Onze uitkomsten suggereren dat het negeren van verschillen in opleiding resulteert in vertekende parameterschattingen. Dat is vooral relevant voor niet-westerse allochtonen die vaak een lager opleidingsniveau hebben. De geschatte loonverschillen tussen autochtonen en niet-westerse allochtonen variëren per toegepast schattingsmodel. Een controle voor de waarnemingen zonder opleidingsinformatie blijkt onvoldoende te zijn om echte loonverschillen tussen groepen te schatten. Dat zou worden veroorzaakt door twee denkbare processen. Ten eerste zou een deel van de migranten hun opleiding in het buitenland hebben gevolgd waardoor de subpopulatie van de waarnemingen zonder opleidinginformatie meer heterogeen wordt. Ten tweede zou het bijbehorende gewicht minder geschikt kunnen zijn voor specifieke migranten groepen..

De eindconclusie is dat het negeren van opleidingsverschillen bij regressie-analyse vanwege de afwezigheid van een geschikte opleidingsvariabele, de schattingen van interetnische verschillen in lonen vertekent. Dat zou nieuwe initiatieven voor onderzoek naar interetnische verschillen in sociaaleconomisch status in de weg kunnen staan. De komst van de SSB-variabele opleidingsniveau kan voorlopig een oplossing betekenen totdat de werknemerspopulatie volledig is terug te vinden in de opleidingsregistraties. Zowel de wijze waarop opleidingregister wordt gebruikt als de interpretatie van de schattingen vragen echter om de nodige zorgvuldigheid.

## *Referenties*

Bakker, Bart F.M., Frank Linder en Dominique van Roon (2008). *Could that be true? Methodological issues when deriving educational attainment from administrative data-sources and surveys* (Shanghai: Paper prepared for the IAOS Conference on Reshaping Official Statistics, 14–16 October 2008)

Friedberg, R.M. (2000). You Can't Take It With You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics*, 18:221–51.

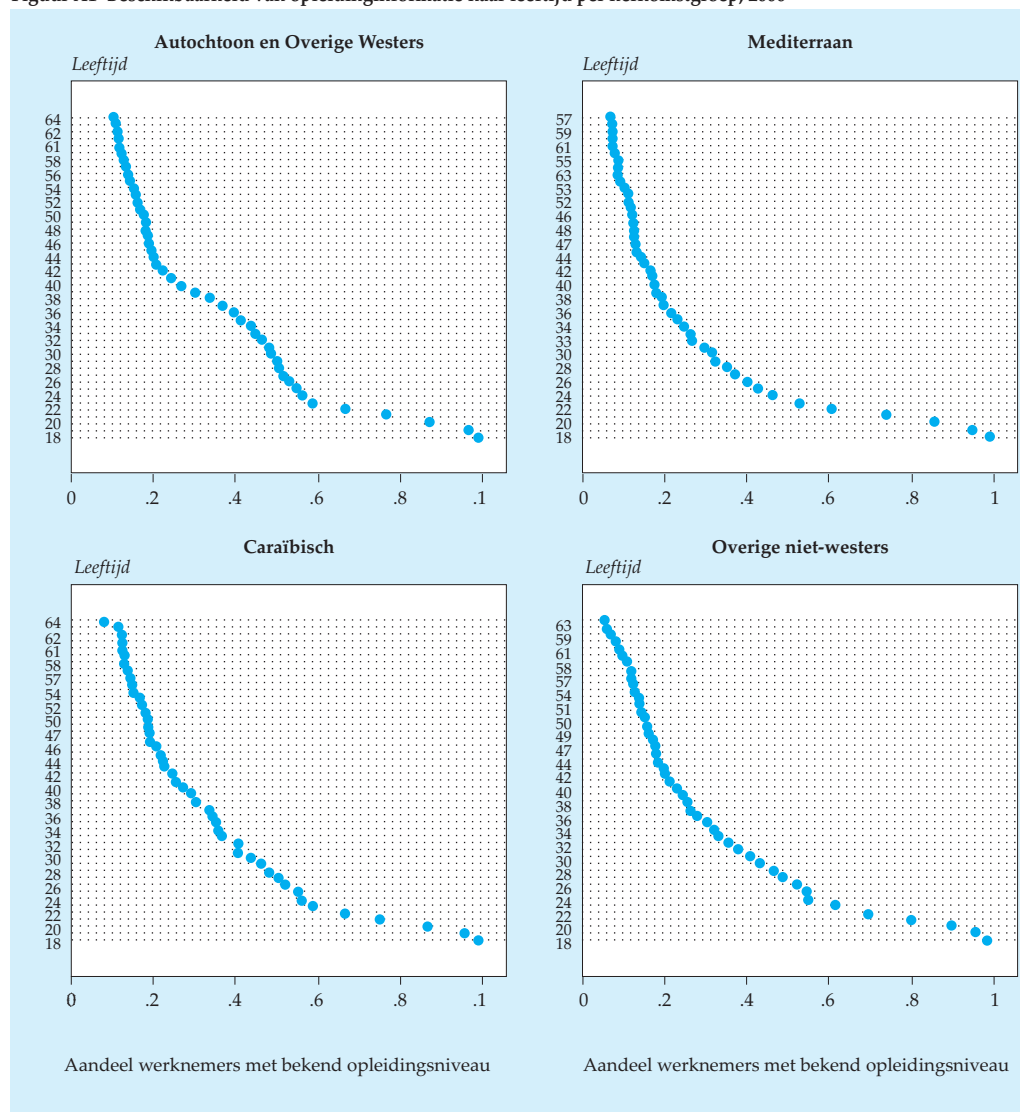
## *Noten*

---

<sup>1)</sup> De figuur voor autochtonen is vrijwel identiek aan de figuur voor tweede generatie westerse migranten (zie ook figuur A1 in appendix).

# Appendix

Figuur A1 Beschikbaarheid van opleidinginformatie naar leeftijd per herkomstgroep, 2006





# Registers en criminologisch onderzoek

Catrien Bijleveld (VU/NSCR)

## 1. *Criminologisch onderzoek*

Criminologen doen onderzoek naar het vóórkomen van criminaliteit (de prevalentie), de etiologie van delinquent gedrag en de (strafrechtelijke) reactie op criminaliteit. Criminologie is een objectwetenschap. Dat wil zeggen dat criminologen hun object van onderzoek gemeen hebben, maar dat de wijze waarop dat object bestudeerd wordt, erg kan verschillen. Verschillen zijn er vooral naar de (moeder)discipline of de wetenschapsopvatting van de onderzoeker. Zo zal een psycholoog andere methoden hanteren dan een antropoloog of een bestuurskundige, en ook op zoek zijn naar andersoortige data. Deze bijdrage is geschreven vanuit een meer kwantitatief perspectief. Deels doordat dat perspectief domineert in het werk van schrijfster dezes, maar vooral omdat registers in dat soort onderzoek bruikbaar zijn. Ik richt mij op gegevens die betrekking hebben op natuurlijke personen als daders, en dus niet op bedrijven. Ook richt ik mij op de dader als object van onderzoek en niet 'feiten', dat wil zeggen misdrijven of overtredingen. Dat wil niet zeggen dat die feiten niet van belang zijn – sterker nog, daar gaat het allemaal om – maar het steekproefkader is de populatie van daders. Ik maak gebruik van ervaringen met longitudinaal onderzoek, onderzoek naar criminele carrières en factoren die daar invloed op hebben.

In veel criminologisch onderzoek wordt gebruik gemaakt van registers. De meest voor de hand liggende zijn de registers die in verband met de opsporing en de strafrechtspleging zijn verzameld. In Nederland is daar – zeker sinds de automatisering in de jaren '90 goed doorzette en vrijwel alle registers werden geautomatiseerd – veel ervaring mee. Nederlandse onderzoekers bevinden zich ten opzichte van onderzoekers in andere landen wat dat betreft in een redelijk luxueuze positie. Allereerst lijkt de kwaliteit van de gegevens in de registers redelijk hoog c.q. het rechtssysteem van Nederland maakt ze goed bruikbaar. Ten tweede krijgen onderzoekers relatief makkelijk toegang. Natuurlijk hebben deze registers ook hun zwakke kanten, redelijk veel overigens. Bij gebrek aan geld en beter doen we het er echter vaak mee.

De laatste jaren wordt in Nederland in toenemende mate gebruik gemaakt van gegevens uit andere registers, dat wil zeggen registers die zijn aangelegd voor andere doeleinden dan verband houdend met criminaliteit. De ervaring daarmee is gering, maar de potentiële winst die te boeken is, is groot.

In deze korte beschouwing geef ik eerst een overzicht van de vaakst gebruikte 'cri-

minologische' registers. Dat zijn registers die zijn aangelegd door instanties die zich bezig houden met opsporing, vervolging, berechting en tenuitvoerlegging. Mijn overzicht zal niet compleet zijn: er zijn nog veel andere registers die (potentieel) bruikbaar zijn. Door de oriëntatie van het onderzoek, onbekendheid of ontoegankelijkheid zijn deze registers mogelijk minder vaak gebruikt. Recente overzichten van dit soort criminologische registers en databronnen zijn te vinden in het periodiek verschijnende *Criminaliteit & Rechtshandhaving* (zie bijvoorbeeld de editie van 2009).

Vervolgens geef ik een korte impressie van het gebruik van twee 'niet-criminologische' registers, te weten de Gemeentelijke Basisadministratie (GBA) en SUWInet. Ik zal proberen te laten zien wat de grote potentie en waarde van deze registers is voor criminologisch onderzoek, maar ook waar de valkuilen zitten. Meer in zijn algemeenheid zal mijn argumentatie er op neer komen dat gezien de populatie die we bestuderen, we meer dan gebruikelijk voorzichtig moeten zijn met het at face value nemen van de geregistreerde werkelijkheid zoals die zich uit deze registers aan ons openbaart. Een deel van de personen die wij bestuderen leeft namelijk levens die er in werkelijkheid mogelijk heel anders uit zien dan de formele werkelijkheid zoals in het register neergelegd.

## 2. *Veel gebruikte criminologische registers*

De politie houdt zich bezig met de opsporing van delicten. Het openbaar ministerie vervolgt een deel van de opgespoorde delinquentie, een nog kleiner deel komt voor de rechter, en het kleinste deel krijgt een gevangenisstraf opgelegd of zit in bewaring of vervangende hechtenis.

De politie registreert sinds 1986 informatie over daders van opgehelderde delicten (dat wil zeggen delicten waarbij een potentiële dader is gevonden) in het Herkenningssysteem ofwel HKS. Ieder korps heeft zijn eigen HKS. Ook heeft de KMar een eigen HKS. Bij het Korps Landelijke PolitieDiensten (KLPD) worden al die lokale HKS databases aan elkaar geplakt en 'ontdubbeld', waarna een analyseerbaar bestand resteert met informatie over zeer veel (170.000 per jaar) unieke verdachten, en de feiten waarvan zij verdacht worden. Dat het hier een database van verdenkingen betreft en niet een database met gegevens over daders is een belangrijke nuancering.

Het HKS bevat informatie over het geboorteland van de verdachte, nationaliteit, sekse, vuurwapengevaarlijkheid en drugsverslaving. Van die laatste moeten we ons realiseren dat die niet longitudinaal beschikbaar zijn: we weten dus in een zeker jaar dat van deze persoon als drugsverslaafd staat geregistreerd, maar wanneer hij of zij daar mee begonnen is is onduidelijk. Per feit zijn er gegevens over pleegplaats, het artikelnummer, of het feit in groepsverband is gepleegd. Het HKS is in de loop der jaren steeds rijker geworden. Het is sinds kort mogelijk om groeps-

daders aan elkaar te koppelen. De database bevat ook gegevens over HALT-afdoeningen die verderop in de strafrechtelijke keten niet meer geregistreerd worden. De database bevat de gegevens over alle feiten waarvan een persoon verdacht is geweest over de periode waarover registraties zijn. Bedacht daarbij dient te worden, dat feiten alleen geschoond worden als een persoon niet als verdachte in het systeem had mogen staan, bijvoorbeeld als hij of zij onterecht als verdachte beschouwd is. Een latere vrijspraak maakt dus niet automatisch dat het feit gewist wordt. De mate waarin HKS daadwerkelijk geschoond wordt, is onduidelijk. Het is dus mogelijk 'verdenkingscarrières' te construeren. Ook bedacht dient te worden, dat de politie feiten mogelijk als zwaarder registreert dan ze uiteindelijk eindigen bij de rechtbank: zo kan iemand ingeschreven worden in HKS voor een zware mishandeling, die bij het OM als mishandeling subsidiair openbare orde wordt vervolgd, waar uiteindelijk een veroordeling voor het laatste uitrolt. Het beeld op basis van HKS is dus mogelijk zwaarder dan het beeld in latere aanleg. HKS bevat nog veel meer 'velden' dan de hier besprokene; de ervaring leert dat die velden vaak zo slecht gevuld zijn (bijvoorbeeld de slachtofferinformatie) dat die in de praktijk niet tot nauwelijks bruikbaar zijn.

Het HKS is daarmee een aantrekkelijke database met veel informatie, over de gehele bekende verdachtenpopulatie van een jaar, zo onder een druk van de knop. Vanzelfsprekend dient toestemming verkregen te worden voor het gebruik van de gegevens.

Het Openbaar Ministerie beheert de database OMdata, die een samenvoeging is van de registraties van op de parketten van ingeschreven strafzaken. OMdata bevat ook de beslissingen op deze zaken in eerste aanleg. In die zin bevat OMdata geen gegevens over personen maar over zaken (waarin zich meerdere feiten kunnen bevinden). Aangezien met behulp van OMdata wel hele populaties van zaken getrokken kunnen worden (bijvoorbeeld alle zaken met tenminste één zedendelict van een bepaald jaar) en daarmee de daaraan te koppelen populatie van vervolgte daders voor dat soort delicten, is de database aantrekkelijk. De praktijk leert dat OMdata nogal eens als eerste stap gebruikt wordt, om vervolgens voor de zo geïdentificeerde personen informatie elders te gaan zoeken (bijvoorbeeld in strafdossiers). Ook hier geldt natuurlijk weer dat eerst toestemming verkregen dient te worden, zowel voor het 'trekken' van OMdata als voor latere analyse van gegevens uit strafdossiers.

Een derde bron is de justitiële documentatie (JD). De JD is een compleet register met informatie over alle personen ter vervolging ingeschreven. Het bevat alle informatie die ook in OMdata zit, alsmede de afdoening van zaken in laatste aanleg. De laatste jaren is steeds meer en steeds preciezere informatie toegevoegd aan de informatie op de JDs. Zo wordt sinds 1995 de pleegdatum geregistreerd (althans 'een' pleegdatum – soms blijkt deze gelijk aan de datum van het proces verbaal); ook bevinden zich steeds meer politieafdoeningen en CJIB afdoeningen op de bladen van de JD. De JD bevat daarmee de 'hardste' gegevens over gepleegde feiten.

De JD is echter niet gelijk aan wat wij in de volksmond noemen 'strafbladen'. Zo bevat de JD ook gegevens over vrijspraken, en over seponeringsbeslissingen. Het is gebruikelijk voor criminologisch onderzoek om vrijspraken en zogenoemde technische sepot (de officier seponeert bijvoorbeeld bij gebrek aan bewijs of omdat de verdachte niet de juiste persoon is) niet mee te tellen; we gaan er dan van uit dat de persoon op de JD deze feiten niet gepleegd heeft, en nemen die feiten niet mee in ons onderzoek naar zijn of haar criminele carrière. Wel tellen we de veroordelingen en de beleidssepots mee, zaken waar de officier het om andere dan vervolgingstechnische redenen niet opportuun acht de vervolging door te zetten. Met die praktijk volgen we bij de beoordeling of we deze feiten tot de criminele carrière van deze dader moeten rekenen het oordeel van de laatste instantie die zich over de feiten boog.

Bij de overzetting tussen 1992 en 1996 van een papieren naar een elektronisch systeem zijn alle registraties van personen die toen 80 jaar of ouder waren vernietigd. Dat betekent dat het systeem in beginsel van alle personen die na 1916 geboren zijn, informatie bevat. Die informatie is niet voor alle jaren even makkelijk toegankelijk. Zo moeten we als we gegevens willen hebben van oudere generaties de zogenoemde 'images' bekijken en van de oudste generaties zelfs filmpjes, broze rolletjes die met de hand op een spoeltje gerold moeten worden, en vervolgens gedraaid door een monitor die in de tijd van de oudste James Bond films heel modern oogde – die overigens de laatste keer dat wij hem gebruikten ontplofte. Dit archief, hoe schijnbaar antiek en onactueel ook, zou het verdienen bewaard te worden voor het nageslacht. Een vierde bron is TULP, het registratiesysteem van de tenuitvoerleggingen van vrijheidsbenemingen. Er zijn separate systemen voor het gevangeniswezen, jeugd en voor de TBS. Hoewel TULP diverse velden bevat die in beginsel interessant zijn voor criminologisch onderzoek, leert de ervaring dat ook hier het systeem meer wordt gebruikt voor steekproeftrekking of meer oppervlakkige beschrijvingen dan voor meer diepgaand gebruik van de data. De kwaliteit van de data is onhelder.

Voor al deze databases geldt dat zij in beginsel vanaf zo rond 1994/1995, het jaar waar de grote automatiseringsoperaties in de strafrechtsketen begonnen, redelijke dekking hebben. OMdata wordt voor de beginjaren nog te onbetrouwbaar geacht. Ieder systeem heeft zijn eigen regime van 'wissen'. Over het algemeen geldt dat naarmate de feiten zwaarder zijn, het langer duurt tot die gewist worden, en dat naarmate personen langer niet meer met nieuwe feiten geregistreerd worden, de oude feiten eerder gewist worden. Feiten en zaken waarop een sepot is gevolgd worden in de JD eerder gewist. Voor de JD geldt, als bron van 'hardste' informatie, dat de registraties van overleden personen ook worden gewist, iets om als onderzoeker erg beducht op te zijn – dat wissen betekent overigens niet dat de informatie er niet meer is, hij zit achter het actieve systeem. We moeten beducht op dit soort schoning zijn, omdat de ervaring leert dat binnen de groepen wat zwaardere daders het weer de meer actieve zijn die een grotere kans hebben te overlijden; behalve onderschatting levert het missen van deze overleden personen dus ook een vertekening op.

Nou vraagt de lezer zich misschien af: doen die criminologen het uitsluitend met door anderen, voor andere doeleinden, geregistreerde informatie? Nee dat is zeker niet zo. Veel onderzoek maakt gebruik van slachtofferenquêtes (al zijn de vraagstellingen dan noodzakelijkerwijs minder op de dader(s) gefocust). Ook gebruiken we daderenquêtes die we 'self-report enquetes' noemen. We weten echter dat die met name voor de wat zwaardere feiten slechter ingevuld worden, en we vermoeden dat bepaalde groepen jongeren onderrapporteren. Ook weten we niet goed of jongens en meisjes mogelijk differentieel rapporteren. Die methoden hebben dus ook hun nadelen. De JD en andere systemen hebben naast het feit dat zij in vergelijking tot enquêtes relatief makkelijk beschikbaar zijn daarnaast dus ook als voordeel (selectiviteit – bijvoorbeeld naar sekse of etniciteit – in de opsporing daargelaten) dat ze redelijk hard en uniform zijn. We kunnen ons verlaten op de rechter of de officier, die als justitiële autoriteit hebben gemeend dat persoon X delict Y gepleegd heeft. Ook hebben zij als voordeel dat ze vrij nauwkeurig hele reeksen misdrijven kunnen weergeven, die door de betreffende personen waarschijnlijk niet meer precies te reconstrueren zouden zijn.

Hun grootste nadeel is echter dat zij maar een fractie van de afhankelijke variabele laten zien. Een grove schatting – uitgaand van een gemiddeld aangiftepercentage van 35% en een ophelderingspercentage van 14% – is dat we ongeveer 5% van de delicten in sommige systemen tegenkomen. Daar is natuurlijk van alles aan te nuanceren en op af te dingen, maar het maakt wel duidelijk dat – voor het soort onderzoek waar ik me nu toe beperk, zoals het kwantitatieve criminele carrière onderzoek – onze afhankelijke variabele vaak gebrekkig is. Dat is een opmerkelijke situatie. We nemen aan dat dat topje van de ijsberg dat we zien wel iets kan laten zien van de processen waar we in geïnteresseerd zijn, en de mechanismen die een rol spelen. Overigens is dat topje nou ook weer niet zo klein, en het hebben van een registratie in de JD geen extreem zeldzame gebeurtenis. Van de Nederlandse mannen tussen de 12 en 32 loopt ongeveer een kwart een registratie op (Van der Geest, Langendoen & Bijleveld, 2006).

### 3. *Gegevens uit andere registers*

In het hiernavolgende zal ik twee registers, GBA en SUWInet, behandelen die in criminologisch onderzoek de laatste jaren zijn gebruikt. In de praktijk wordt vaak enigszins gekscherend gezegd dat om criminelen weer op het rechte pad te helpen de drie w's moeten worden ingezet: werk, wijf en woning (deze wijsheid is overigens hoogstwaarschijnlijk slechts op waarnemingen bij mannelijke daders gebaseerd). Informatie over die drie w's bevindt zich voor een deel in deze twee registers. Het zijn zoals boven al gezegd niet de enige interessante registers, maar er is enige ervaring mee en ze zijn bruikbaar om een aantal belangrijke criminologische theorieën mee te toetsen. Bij dat soort theorieën gaat het dan bijvoorbeeld om de

vraag of het hebben van een partner de kans op crimineel gedrag vermindert. Veel onderzoek wijst daar op. Wat nog wel eens wordt bediscussieerd, is hoe dat dan komt. Is het de sociale controle van de partner? Het wegvallen van het netwerk van deviante vrienden? Vinden we het effect voor alle huwelijksrelaties, of alleen voor relaties die als 'goed' worden bestempeld? Overigens wordt voor het hebben van kinderen wel gevonden dat het risico op delinquentie juist zou verhogen.

Soortgelijke vervolgvragen kan men ook stellen voor het veronderstelde dempende effect van werk. Komt dat doordat men tijd doorbrengt met meer prosociale anderen die andere normen hebben? Komt het omdat men geld heeft, en dus niet meer hoeft te stelen om in allerlei behoeften te voorzien? Komt het omdat men met werken (wat dan sterker zou gelden voor vast werk) een zeker sociaal 'kapitaal' opbouwt dat men niet graag opgeeft voor delinquentie?

Het eerste register is de GBA. In de GBA worden gegevens met betrekking tot de burgerlijke stand en woonplaats geadministreerd. Het is sinds 1994 operationeel, toen de informatie van de tot dan gebruikte zogenoemde persoonskaarten is overgezet naar een elektronisch systeem. Informatie van alle toen levende personen is overgezet; niet alle gemeenten hebben ook alle oudere informatie van deze personen overgezet. Als we onderzoek doen over personen die voor 1994 overleden zijn, moeten we terug naar de papieren persoonskaarten, waarvan bij het Centraal Bureau voor de Genealogie kopieën te krijgen zijn. De GBA bevat informatie over adressen, de namen van de ouders, huwelijk, echtscheiding, en kinderen. Er bevindt zich ook informatie over ondercuratelestelling en toezichtrelaties. Onderzoekers krijgen doorgaans niet voor alle velden toestemming. Zo kregen wij in eerder onderzoek geen toestemming om de variabele geslacht te zien.

Het tweede register is SUWInet. SUWInet is een register dat gebruikt wordt door allerlei instanties die te maken hebben met werkzoekenden en uitkeringsgerechtigden. SUWInet bevat gegevens afkomstig van CWI, de Sociale Diensten, UWV, en GBA. In het register staan op persoonsniveau, ook historisch, gegevens over inschrijvingen bij CWI. De gegevens van de sociale diensten hebben betrekking op uitkeringen en uitkeringsaanvragen in het kader van de WWB. De gegevens van UWV hebben betrekking op arbeidsrelaties en uitkeringsverhoudingen. Van personen overleden voor 1992 is niet waarschijnlijk dat er informatie geregistreerd is. Van daarna overleden personen is wel historische informatie te vinden. Men kan bijvoorbeeld bij een aanstelling die een persoon heeft gehad dóórzoeken om te zien waar het werkverband was. Niet altijd is dan te achterhalen wat de persoon voor werk heeft gedaan, slechts dat hij of zij bij firma X heeft gewerkt. Ook is niet altijd te achterhalen voor hoeveel dagen per week iemand heeft gewerkt: we kunnen slechts zien dat er een aanstelling was bij uitzendbureau Y van 11 maart tot 11 oktober, maar de deeltijdfactor wordt niet aangegeven. Bij uitkeringen kunnen we zien of iemand bijvoorbeeld een daklozenuitkering heeft gehad.

Voor deze bijdrage heb ik informatie uit diverse andere bronnen over de levenslopen van mannen die een hoog-risico lopen op delinquentie gelegd naast gegevens

uit de GBA en gegevens uit SUWInet. Het betreft een dataset (Van der Geest & Bijleveld, 2008) met gegevens van 270 mannen die in hun jeugd voor ernstige gedragsproblemen behandeld zijn. Het gaat dus om een groep 'hoog-risico mannen'. Na behandeling recidiveerde 85%, 76% naar een ernstig delict. Meer dan de helft werd na het 18e jaar ooit gedetineerd. Ongeveer 30% trouwde; 40% kreeg kinderen. Tegen de 60% werkte tenminste één dag in de legale economie.

Van sommigen hebben wij nadere informatie over hun levenslopen vergeleken met hun levenslopen zoals uit GBA en SUWInet naar voren komt. Van 11 mannen hadden we nadere gegevens die meerdere perioden van hun leven bestreek en ook redelijk precies was in de zin dat werd aangegeven wanneer er gewerkt was, wanneer kinderen geboren waren, of waar iemand gewoond had. Deze mannen vormen waarschijnlijk geen afspiegeling van de groep van 270 hoog-risico mannen. De hoog-risico mannen vormen eveneens geen afspiegeling van delinquente mannen in Nederland. Daarmee dient het hiernavolgende als met name illustratief te worden opgevat. Ik behandel de bevindingen per relevant domein.

#### 4. *Relaties en kinderen*

Twee mannen waren getrouwd geweest; een van deze twee huwelijken vonden we niet in de GBA terug. Zes van de 11 mannen waren nooit getrouwd geweest, maar hadden wel samengewoond of een vriendin. Deze samenwoningsrelaties werden niet in de GBA teruggevonden. Dat is ook niet verwonderlijk; alleen bij een officiële verbintenis verwachten we die als zodanig terug te vinden. Op zich zou deze informatie technisch gesproken wel bij benadering te vinden moeten zijn: men kan van betrokkenen nagaan of op het woonadres slechts een persoon van de andere sekse woont, en dan aannemen dat deze twee personen een samenwoningsrelatie hebben. Dat is natuurlijk niet waterdicht. En het is ook een nogal omslachtige nazoeckklus. Men heeft daar ook een ruimere toestemming voor nodig. Al met al betekent dit dat wij voor deze mannen onderschatten hoe vaak zij samenwonen met een partner.

Vijf van de 11 mannen hadden één of meer kinderen. Op een kind na, werden alle kinderen teruggevonden in de GBA; bijna alle mannen hadden dus de door hen verwekte kinderen die zij kenden wettelijk erkend. Het gaat hier natuurlijk wel om kleine aantallen.

#### 5. *Werk*

De gegevens over werk leken redelijk goed te kloppen met SUWInet. Vooral oudere aanstellingen werden echter nog wel eens niet teruggevonden. Dat kan mogelijk ook gelegen hebben aan het feit dat men zwart gewerkt heeft, maar dat dat in de verslag-

leggingen die wij inzagen als wit werk is geïnterpreteerd. Regelmatig gaf men ook aan zwart gewerkt te hebben, of met de kermis mee te reizen. We vonden zelfs ook dat mensen verklaarden zwart voor een uitzendbureau gewerkt te hebben.

Buiten de 11 meer in detail onder de loep genomen mannen hield een deel van de 270 mannen zich soms ook in leven met stelen en helen. Een deel was niet in staat een uitkering aan te vragen, een deel kreeg van de ouders of een vriendin geld toegestopt waar hij dan van leefde.

SUWInet bevat (recent) ook informatie over opleidingen. Van deze opleiding is echter niet goed te zien wanneer het diploma behaald is, en daarmee is die lastig aan de criminele carrière te koppelen.

Hoewel de deelname aan de legale economie dus redelijk goed terug lijkt te komen, lijkt voor deze groep mannen het zwarte circuit van groot belang om in hun levensonderhoud te voorzien. Hoewel we met de SUWInet gegevens dus wel bijvoorbeeld theorieën over sociaal kapitaal kunnen toetsen, kunnen we voor deze populatie niet goed zien in hoeverre verklaringen die zich meer richten op geldbehoefte opgeld doen.

## 6. *Wonen*

Op het gebied van wonen zagen we een aantal opmerkelijke afwijkingen tussen de GBA en onze overige informatie. Bij drie mannen lazen we dat ze weliswaar op een bepaald adres ingeschreven stonden maar in werkelijkheid dan weer bij vriend X en dan weer bij vriendin Y verbleven, dat ze bij moeder ingeschreven stonden maar afwisselend bij een vriendin of op de camping woonden. Een man woonde jarenlang overal en nergens – maar stond heel stabiel al die jaren op een en hetzelfde adres ingeschreven. Van een man klopte van zes adressen in de GBA er slechts een, in die zin dat deze persoon daar toen ook daadwerkelijk woonde. De formele woonwerkelijkheid zou al met al wel eens relatief het meest kunnen afwijken van het daadwerkelijke woonadres. Voor zwerven gaf de GBA overigens meestal wel een vrij goede indicatie (mannen hadden dan een briefadres of stonden ingeschreven bij instellingen voor dagopvang).

## 7. *Conclusie*

Criminologen kunnen veel mooie dingen doen met registers. Criminologische registers lijken vooral voor onderzoek dat een kortere tijdspanne bestrijkt, en niet lang terug hoeft te kijken, een grote potentie te hebben. Dat is niet alleen de relatief eenvoudige beschikbaarheid van de data, maar ook de precisie waarmee aanstellingen, detenties, huwelijken en delinquent gedrag aan elkaar gekoppeld kunnen worden. Nederlandse onderzoekers hebben, vooral door hun toegang tot criminologische registers, een grote voorsprong op hun buitenlandse collega's. Onderzoek



dat langer terugkijkt, of vanuit een eerder startpunt dan het ontstaan van de grote elektronische registers vooruit kijkt, zal vaak moeten teruggrijpen op andere bronnen, deels niet-elektronische registers.

Criminologische registers hebben hun beperkingen, waar we voor een deel voldoende zicht op hebben. Voor niet-criminologische registers is dat nog lang niet het geval. De groep mannen voor wie ik dat hier onderzocht heb is in die zin een extreme keuze. Voor hoog-risico vrouwen zijn de afwijkingen mogelijk al geringer. Niettemin zou het aanbeveling verdienen om op meer systematische wijze te verkennen in hoeverre de elektronische werkelijkheid van registers de leefsituatie van in criminologisch onderzoek voorkomende soms deviante populaties weerspiegelt. Hoewel het denkbaar is dat voor sommige vragen (bijvoorbeeld deelname aan de legale arbeidsmarkt) het elektronisch register alle informatie heeft die men wenst – beter en preciezer dan ooit (retrospectief) reproduceerbaar door respondenten zelf – is het klakkeloos voor waar aannemen van de register-werkelijkheid riskant. Een meer systematische verkenning is nodig om te weten waar we de geringste en zwaarste vertekeningen mogen verwachten, en in welke richting.

## *Referenties*

S.N. Kalidien, S.N. & Eggen, A.Th.J. (Red). (2009). *Criminaliteit en rechtshandhaving 2008. Ontwikkelingen en samenhangen*. Reeks Onderzoek en Beleid, 279. den Haag: Boom Juridische Uitgevers.

Van der Geest, V., Langendoen, M. & Bijleveld, C.C.J.H. (2006). Brave burgers? IJkingsonderzoek naar justitiecontacten van Nederlandse mannen. *Secondant*, 20 (december), 31–35.

Van der Geest, V. & Bijleveld, C.C.J.H. (2008). Personal, background and treatment characteristics associated with offending after residential treatment: A 13-year follow up in adolescent males. *Psychology, Crime & Law*, 14, 159–176.



# *Nauwkeurigheid van gerapporteerde giften aan goededoelenorganisaties*

René Bekkers (VU)<sup>1)</sup>

## *1. Het probleem van onnauwkeurigheid*

Vanzelfsprekend is de waarde van onderzoek groter naarmate het nauwkeuriger schattingen oplevert. Dit geldt ook voor onderzoek op basis van enquêtes, dat in de sociale wetenschappen in de afgelopen decennia de standaard is geworden. Een probleem dat al snel optreedt wanneer respondenten in enquêtes niet nauwkeurig rapporteren is dat generalisatie naar de doelpopulatie systematische fouten bevat. Een bekende bron van fouten betreft de steekproefsamenstelling: wanneer de steekproef onevenredig veel respondenten bevat die systematisch te hoog (of te laag) scoren op het verschijnsel waarvan de populatiewaarde geschat moet worden, zal die schatting systematisch te hoog (of te laag) uitvallen.

Veel onderzoek naar nauwkeurigheid van rapportages gaat over onwenselijk gedrag, zoals seksueel misbruik en kindermishandeling (Hardt & Rutter, 2004), het gebruik van alcohol (Midanik, 1982) en drugs (Harrell, 1997) en over criminaliteit (Thornberry & Krohn, 2004). Er is ook veel onderzoek naar gezondheidsklachten (zie bijvoorbeeld Hahn et al., 1997). In deze onderzoeksgebieden is een nauwkeurige rapportage van belang voor de optimale behandeling van slachtoffers en/of daders. Minder onderzoek is er gedaan naar de nauwkeurigheid van sociaalwenselijk gedrag, zoals het stemmen bij verkiezingen. Dit is begrijpelijk omdat er door ongeldige conclusies over sociaalwenselijk gedrag als gevolg van onnauwkeurigheid geen mensen overlijden en er geen hoge kosten mee gemoeid zijn.

Maar er is nog een andere reden om bezorgd te zijn over onnauwkeurigheid in rapportages door respondenten in enquêtes. Als enquêtegegevens gebruikt worden om theorieën te toetsen over verbanden tussen verschillende sociale verschijnselen kunnen onnauwkeurige rapportages tot ongeldige conclusies over deze verbanden leiden. Deze conclusies kunnen net als de rapportages systematische fouten bevatten. In het onderzoek naar stemgedrag is gesteld dat het gebruik van zelfrapportages 'common sense' theorieën over de samenleving ondersteunt (Presser & Traugott, 1992).

De algemene veronderstelling in de literatuur over de effecten van onnauwkeurigheid op de geldigheid van conclusies luidt dat vertekening in zelfrapportages problematischer is naarmate de vertekening een sterkere samenhang vertoont met het verschijnsel dat wordt bestudeerd (Katosh & Traugott, 1981; Presser, 1984;

Phillips & Clancy, 1972; Bernstein, Chadha, & Montjoy, 2001; Presser & Traugott, 1992). Een voorbeeld op het terrein waarop ook de analyses later in dit artikel zich begeven maakt deze veronderstelling duidelijk. De hoogte van bestedingen door huishoudens neemt toe met het inkomen. Maar wanneer leden van huishoudens met een hoger inkomen in enquêtes minder nauwkeurig zijn in hun rapportage van bestedingen en de nauwkeurigheid afneemt naarmate het inkomen stijgt, zal het verband tussen het huishoudinkomen en de uitgaven onderschat worden. In ons geval zou het verband tussen het inkomen en giften aan goededoelenorganisaties op basis van enquêtes zwakker lijken dan het in werkelijkheid is.

## 2. *Bronnen van vertekening*

De wijze waarop respondenten in enquêtes rapporteren over verschijnselen waarover zij informatie hebben kan op een vruchtbare manier worden bestudeerd vanuit sociaalpsychologische theorieën over sociaal gedrag (Tourangeau, Rips & Rasinski, 2000). In deze theorieën wordt menselijk gedrag gezien als een resultaat van het nastreven van doelen, gegeven de moeite die dat kost en de mogelijkheden die de situatie daarvoor biedt. In deze bijdrage ontbreekt de ruimte om een algemeen theoretisch kader voor rapportages in enquêtes verder uit te werken. Ik volsta daarom met een kort overzicht van de literatuur op de meest relevante onderzoeksgebieden.

Respondenten geven in een interviewsituatie meer nauwkeurige antwoorden naarmate zij daartoe sterker gemotiveerd zijn, bijvoorbeeld omdat zij het onderwerp interessanter vinden. Respondenten kunnen in de interviewsituatie echter ook andere doelen nastreven. Respondenten vinden het ook belangrijk dat zij door anderen positief beoordeeld worden en stemmen hun antwoorden af op wat zij verwachten dat anderen positief waarderen (wat de sociale norm is). Tenslotte vinden respondenten het prettig een positief beeld van zichzelf te geven omdat dit bijdraagt aan hun psychologisch welbevinden.

Respondenten geven in een enquête meer nauwkeurige antwoorden naarmate de interviewsituatie dat gemakkelijker maakt. Wanneer respondenten door de manier waarop de vragen gesteld worden meer geholpen worden nauwkeurige antwoorden te geven zullen zij nauwkeuriger zijn. De doelen die respondenten in interviewsituaties hebben komen in veel gevallen tot uiting in sociaalwenselijke antwoorden: zij doen zichzelf bijvoorbeeld voor als minder crimineel, gezonder, sterker betrokken bij de politiek en vrijgevinger dan zij in werkelijkheid zijn. Sociaalwenselijke antwoorden worden daardoor minder vaak gegeven naarmate het onderwerp van de enquête meer normatief geladen is (Tourangeau, Rips & Rasinski, 2000).

Sociaalwenselijke antwoorden worden eveneens minder vaak gegeven wanneer de respondenten die niet voldoen aan de norm zonder gezichtsverlies een antwoord-

categorie kunnen kiezen waardoor ze niet hoeven te liegen (Belli, Traugott, Young & McGonagle, 1999), en er geen mensen bij het invullen van de enquête aanwezig zijn die sociaalwenselijke antwoorden kunnen goedkeuren (Tourangeau & Smith, 1996; Tourangeau, Jobe, Pratt, and Rasinski, 1997; Stocké, 2007a). Ook geven respondenten vaker nauwkeurige antwoorden wanneer het om saillante gebeurtenissen gaat (Sudman & Bradburn, 1973).

Naar aanleiding van het klassieke werk van Crowne & Marlowe (1960, 1964) is er veel onderzoek gedaan naar de mogelijkheid dat de tendens om sociaalwenselijke antwoorden te geven een stabiele eigenschap van respondenten is. De consensus is intussen echter dat het geven van sociaalwenselijke antwoorden sterk afhankelijk is van de situatie. Hoewel sociaalwenselijk gedrag wel een zekere mate van consistentie vertoont (zie bijvoorbeeld Wilhelm & Bekkers, 2010), hangt het geven van sociaalwenselijke antwoorden op verschillende soorten vragen nauwelijks met elkaar samen (Presser, 1984; Clark & Tiff, 1966). De conventionele meetinstrumenten voor sociale wenselijkheid zijn eerder een maat voor daadwerkelijk sociaalwenselijk gedrag (Barger, 2002; McCrae & Costa, 1983; Mills & Kroner, 2005) en hangen niet samen met de nauwkeurigheid van rapportages in enquêtes (Johnson, O'Rourke, Burris & Warnecke, 2005). Overigens bleek in een Nederlandse studie waarin giften aan KWF Kankerbestrijding zoals gerapporteerd in een enquête werden vergeleken met de organisatie geregistreerde giften dat respondenten die het vaker eens waren met de stelling 'Ik doe alles om het anderen naar de zin te maken' hun giften sterker overdreven (Bekkers & Wiepking, 2010).

De zoektocht naar andere kenmerken van respondenten die samenhangen met de kans op sociaalwenselijke antwoorden is tot nu toe eveneens vrij onvruchtbaar gebleken (Presser & Stinson, 1998). Sommige studies vinden dat het opleidingsniveau positief samenhangt met de nauwkeurigheid van rapportages. Dit betreft gezondheidsgelateerde verschijnselen (zie bijvoorbeeld Hahn et al., 1997) en de recentere studies naar stemgedrag (Silver, Anderson & Abrahamson, 1996). In studies naar stemgedrag van voor 1980 wordt nauwelijks een verband met het opleidingsniveau aangetroffen (Kathosh & Traugott, 1981). In de klassieke Denver-validatiestudie uit de jaren '40 werd een positief verband gevonden tussen het opleidingsniveau en de nauwkeurigheid van gerapporteerd geefgedrag (Cahalan 1968). In de meer recente KWF-validatiestudie werd echter een negatief verband gevonden (Bekkers & Wiepking, 2010): hoger opgeleiden overdreven hun giften meer dan lager opgeleiden.

In het empirische gedeelte van dit artikel onderzoek ik de nauwkeurigheid van door respondenten in een enquête gerapporteerde giften. Onderzoekers in het vakgebied van filantropische studies (voor een overzicht van de literatuur zie Bekkers & Wiepking, 2007) gebruiken in de regel enquêtes om giften vast te stellen bij gebrek aan registerdata op dit gebied. Sinds het midden van de jaren '90 wordt het volume en de aard van de filantropische geldstromen in Nederland vastgesteld met gegevens uit 'Geven in Nederland' enquêtes die om het jaar georganiseerd worden door de Vrije Universiteit. Op basis van deze gegevens

wordt geschat dat het totale volume aan donaties aan goededoelenorganisaties in Nederland in 2007 € 1 945 miljard euro bedraagt (Schuyt, Gouwenberg & Bekkers, 2009).

Twee studies zijn in het buitenland verricht naar de nauwkeurigheid van gerapporteerde giften. In de Denver-validiteitstudie bleek dat 34% van de gerapporteerde giften niet daadwerkelijk waren ontvangen (Parry & Crossley, 1950). Burt en Popple (1998) vonden in een onderzoek onder studenten in Australië een overschatting van giften met 30% wanneer de studenten niet wisten dat hun antwoorden gecontroleerd zouden worden. Wanneer ze dat wel wisten was de overschatting veel kleiner.

In Nederland is eerder met twee registraties nagegaan wat de nauwkeurigheid is van rapportages in enquêtes over geefgedrag. In beide studies betreft het enquêtegegevens uit het Geven in Nederland-onderzoek. Een vergelijking van giften aan de R.K.-Kerk op macro-niveau heeft laten zien dat in het Geven in Nederland-onderzoek de giften met zo'n 22% worden onderschat (Schuyt, Gouwenberg & Bekkers, 2009, p. 206). Een vergelijking van giften aan KWF Kankerbestrijding op individueel niveau heeft laten zien dat de giften met ongeveer 30% werden overdreven door respondenten die in werkelijkheid giften hebben gedaan en die ook hadden gerapporteerd (*true positives*). *False positives* (gerapporteerde giften die in werkelijkheid niet zijn gedaan) en *false negatives* (giften die zijn gedaan maar niet zijn gerapporteerd) konden door de opzet van het onderzoek niet worden waargenomen.

### 3. *Data en methoden*

#### 3.1 *Schatting van vertekening*

Om de mate van vertekening te schatten is een vergelijking van diverse bronnen van gegevens nodig. In dit artikel vergelijk ik de giften aan Greenpeace Nederland zoals gerapporteerd door respondenten in een enquête met de giften afkomstig van de huisadressen waarop deze respondenten woonachtig zijn zoals die zijn geregistreerd door Greenpeace. Uitgedrukt in de totale baten uit eigen fondsenwerving is Greenpeace een van de tien grootste goededoelenorganisaties in Nederland (CBF, 2009). Deze organisatie is gekozen omdat alle giften ook geregistreerd worden in een database. Een probleem in de eerdere KWF-validatiestudie (Bekkers & Wiepking, 2010) is dat niet alle giften aan KWF ook geregistreerd worden, waardoor het aantal *false positives* en de mate van vertekening niet kan worden vastgesteld. KWF krijgt veel kleine giften binnen via de huis-aan-huis-collecte. Deze giften worden in de regel niet afzonderlijk op huisadres geregistreerd. Greenpeace houdt geen huis-aan-huis-collecte maar krijgt uitsluitend giften per bank/giro binnen. Van de giften is daardoor wel bekend van welk huisadres ze afkomstig zijn. Hierdoor kunnen ook *false positives* worden vastgesteld.

### 3.2 *Enquêtegegevens*

De enquêtegegevens die in dit artikel worden gebruikt zijn afkomstig uit het Nederlands Donor Panel (NDP), een online marktonderzoek naar vertrouwen in goededoelenorganisaties en geefgedrag dat elk kwartaal wordt gehouden door WWAV, een bureau voor non-profit marketing en communicatie (WWAV, 2009). NDP-respondenten worden geworven via e-mail uit de e-mailadressen van consumenten die eerder hebben deelgenomen aan de 'Grote Consumenten Enquête' (GCE). De GCE is een tweejaarlijks *mixed-mode* consumentenonderzoek naar de bestedingen en voorkeuren van huishoudens. De vragenlijst wordt schriftelijk verspreid onder alle huishoudens in Nederland, maar kan ook *online* worden ingevuld. Aan elke editie van het NDP nemen zo'n 200 000 respondenten deel. Uitnodigingen voor het NDP worden gestuurd naar toevallige selecties van GCE-respondenten van wie een geldig e-mail adres bekend is.

In de hierna volgende analyses maak ik gebruik van de gegevens uit de NDP editie van mei 2007. In totaal werden uitnodigingen om 'deel te nemen aan een onderzoek over goede doelen' verstuurd naar 60 000 e-mailadressen. Zo'n 2 300 respondenten vulden de enquête volledig in. De respons van 3,8% is laag vergeleken met typische respons percentages voor persoonlijke interviews of schriftelijke vragenlijsten die via reguliere post worden verstuurd (Stoop, 2005). Ook is de respons niet representatief voor de gehele Nederlandse bevolking. De respondenten waren vaker vrouw (57%), hadden minder vaak betaald werk (27% tegenover 74% volgens CBS Statline), hadden minder vaak een HBO- of universitair diploma (8% tegenover 24% volgens CBS Statline), en zeiden vaker dat ze bloed hadden gegeven (8% tegenover 3%) en gestemd hadden bij de afgelopen Tweede Kamerverkiezingen (88% tegenover 80%). Het doel van dit artikel is echter om meetfouten te schatten, en niet om steekproeffouten te onderzoeken. Het is de vraag of de respondenten in de enquête nauwkeurig hebben gerapporteerd, en niet of de steekproef representatief is. Bij de interpretatie van de resultaten is het wel van belang om de samenstelling van de steekproef in het achterhoofd te houden. Het is mogelijk dat juist de meer nauwkeurige respondenten vaker aan het onderzoek hebben meegedaan.

In de NDP-enquête zijn vragen opgenomen over giften aan goededoelenorganisaties, die zijn overgenomen uit de 'Geven in Nederland' (GIN) vragenlijst (Bekkers & Wiepking, 2010). In de NDP-enquête zeiden 1,463 respondenten (64%) dat (iemand in) het huishouden in het afgelopen jaar geld heeft gegeven aan goededoelenorganisaties via bank/giro of internet. Deze respondenten kregen vervolgens een lijst te zien met organisaties met de vraag om aan te geven aan welke van deze organisaties (iemand in) het huishouden had gegeven. Wanneer respondenten een organisatie aankruisten kregen zij de vervolgvraag hoeveel ze hadden gegeven in het afgelopen jaar. In de NDP-enquête werd in tegenstelling tot in het GIN-onderzoek geen 'weet niet' antwoordcategorie aangeboden bij de vraag naar het bedrag. Omdat 'weet niet' een antwoord is waarmee respondenten die onterecht hebben beweerd gegeven te hebben niet hoeven te liegen kan de mate van sociale wense-

lijkheid in het NDP hoger zijn dan in GIN. Het NDP is net als GIN een *online* vragenlijst, waardoor sociaalwenselijke antwoorden verminderd zouden moeten worden omdat er geen interviewer aanwezig is.

### 3.3 *Koppeling*

In beide datasets waren gegevens opgenomen over huisadressen in de vorm van de postcode (4 cijfers en 2 letters) en het huisnummer. Aan Greenpeace Nederland werden alle postcodehuisnummercombinaties van respondenten in het NDP ter beschikking gesteld. De datamanager van Greenpeace verrichte een koppeling op basis van de unieke postcodehuisnummercombinaties, die 322 geslaagde koppelingen opleverde. Greenpeace zette alle giften vanuit deze adressen uit de periode van 1 mei 2006 tot en met 31 mei 2007 uit de database in een apart bestand, dat de auteur vervolgens heeft gekoppeld aan de NDP-data.

## 4. *Resultaten*

### 4.1 *Totale nauwkeurigheid*

In tabel 1 is de vertekening te zien in de rapportages over de giften door de NDP-respondenten. Door 85% van de respondenten die aan de enquête deelnamen is geen gift gerapporteerd en ook geen gift gedaan. Van iets meer dan 7% van de respondenten werd een gift ontvangen die bovendien werd gerapporteerd. Iets meer dan 4% van de respondenten rapporteert giften die niet zijn ontvangen (*false positives*). Van iets minder dan 4% zijn wel giften ontvangen die niet zijn gerapporteerd. Dit betekent dat 35% van alle geregistreerde giften niet werd gerapporteerd door de respondenten. Van de giften die werden gerapporteerd is 38% niet geregistreerd.

Opvallend genoeg rapporteren respondenten van wie geen giften zijn geregistreerd (*false positives*) lagere giften (€ 21,01) dan respondenten die giften rapporteren die wel in de database zijn opgenomen (*true positives*, € 28,22). Dit verschil is sterk significant ( $F=10\ 471$ ,  $p<.001$ ). Met andere woorden: als respondenten giften rapporteren die in werkelijkheid niet zijn gedaan zijn dat relatief kleine leugens, wellicht als gevolg van de behoefte om sociaalwenselijk te antwoorden (Presser & Traugott, 1992).

Donaties die wel zijn geregistreerd maar niet gerapporteerd werden (*false negatives*: € 25,68) zijn wat lager dan de *true positives*. Dit verschil ondersteunt de saillantie hypothese die voorspelt dat kleinere giften gemakkelijker worden vergeten (Bekkers & Wiepking, 2010). Het verschil is echter niet erg groot en niet significant ( $F=1\ 773$ ,  $p<.183$ ). Onder de respondenten die daadwerkelijk giften hadden gedaan en dat ook gerapporteerd hadden waren de geregistreerde giften zo'n 14% lager dan de gerapporteerde giften. Dit verschil is significant in een gepaarde t-toets ( $t=2\ 181$ ,  $p<.031$ ).



**Tabel 1**  
Vertekening in gerapporteerde giften

	Gift niet gerapporteerd	Gift wel gerapporteerd
Gift niet geregistreerd	'True negatives' 84,6% (n=1945)	'False positives' 4,4% € 21,01 gerapporteerd (n=101)
Gift wel geregistreerd	'False negatives' 3,9% (n=89) € 25,68 geregistreerd	'True positives' 7,2% € 32,88 gerapporteerd (n=165) € 28,22 geregistreerd

Wanneer alle positieve waarnemingen van zowel gerapporteerde als geregistreerde giften worden gebruikt (alle gevallen behalve de *true negatives*) blijkt het grootste negatieve verschil tussen gerapporteerde en geregistreerde giften een onderrapportage van € 145 te zijn. Het grootste positieve verschil is een overrapportage van € 155. Het gemiddelde verschil is € 0,26. De mediaan en de modus van het verschil tussen gerapporteerde en geregistreerde giften zijn allebei € 0.

#### 4.2 *Wie overdrijft het geefgedrag?*

Welke kenmerken van respondenten gaan samen met verschillen tussen de gerapporteerde en geregistreerde giften? Er werden drie groepen onderscheiden van ongeveer gelijke omvang: respondenten die giften rapporteerden die meer dan € 5 onder de geregistreerde giften lagen (de onderrapporteerders; 33,2%), respondenten die giften rapporteerden die meer dan € 5 boven de geregistreerde giften lagen (de overrapporteerders, 32,1%), en de respondenten die een bedrag rapporteerden dat minder dan € 5 van het geregistreerde bedrag lag (33,5%). Zijn onderrapporteerders het spiegelbeeld van overrapporteerders of zijn zij een wezenlijk andere categorie van respondenten, met geheel eigen kenmerken? Deze vraag wordt beantwoord in tabel 2, waarin een multinomiale logistische regressie-analyse van de richting van vertekening wordt gepresenteerd. De referentiecategorie wordt gevormd door de respondenten die het bedrag ongeveer goed (+/- € 5) hadden gerapporteerd. De resultaten in tabel 2 laten zien dat de kenmerken van overrapporteerders en onderrapporteerders nogal verschillend zijn. Voor zes kenmerken geldt dat het teken van de variabele in de regressievergelijking verschilt tussen de twee contrasten. Voor geen enkele van deze kenmerken zijn de parameters echter significant. voor zeven variabelen is het teken wel hetzelfde in beide contrasten. Twee hiervan zijn significant (in beide contrasten). Dit zijn het stemmen bij de vorige verkiezingen en het vertrouwen in goedbedoelenorganisaties.

Respondenten die ouder zijn dan 65 of van wie de leeftijd onbekend is rapporteren vaker lager dan de geregistreerde bedragen, terwijl tweeverdieners juist vaker het bedrag goed hebben in plaats van onderrapporteren. Kerkgangers en responden-

**Tabel 2**  
**Multinomiale logistische regressie van de richting van vertekening in gerapporteerde giften**

	onderrapportage		overrapportage	
	<i>odds ratio</i>			
Vrouw	0,90	0,89	1,21	1,18
Ouder dan 65	2,15	2,51	1,28	1,48
Leeftijd onbekend	2,54*	2,64*	1,02	1,13
Tweeverdiener	0,51(*)	0,55(*)	1,02	1,09
Geen betaald werk	0,62	0,54	0,91	0,75
Opleidingsniveau	1,06	1,11	1,08	1,13
Partner in huishouden	0,69	0,74	1,67	1,71
Inkomen (log)	0,71	0,74	1,47	1,10
Eigen huis	1,33	1,35	0,90	0,92
Kerkgang	0,76	0,84	2,02*	2,30*
Gestemd bij laatste verkiezingen		0,24*		0,30(*)
Bloeddonor		0,70		0,58
Vrijwilliger		1,21		0,74
Totaal gegeven bedrag (log)''		1,12		1,80***
Vertrouwen in goededoelenorganisaties		0,58**		0,63*

\*\*\* p<.001; \*\* p<.01; \* p<.05; (\*) p<.10.

Referentie categorie: rapportage minder dan € 5 van geregistreerde bedrag.

ten die ook hogere giften aan andere goededoelenorganisaties rapporteerden gaven vaker te hoge giften aan Greenpeace op. Respondenten die rapporteren dat ze hebben gestemd bij de afgelopen verkiezingen en respondenten die een hoger vertrouwen in goededoelenorganisaties zeggen te hebben rapporteren vaker hun giften accuraat in plaats van dat ze de giften over- of onderdrijven. Dit geldt ook voor respondenten die zeggen geen betaald werk te hebben en respondenten die zeggen bloed te geven, hoewel deze variabelen de conventionele significantieniveaus (net) niet bereiken. Opvallend is dat geslacht, het opleidingsniveau, inkomen en eigen huisbezit geen significant verband vertonen met de richting van vertekening. In tabel 3 wordt de analyse herhaald voor de hoogte van het verschil tussen gerapporteerde en geregistreerde giften. De afhankelijke variabele in deze analyse is het gerapporteerde bedrag verminderd met het geregistreerde bedrag. Hoe hoger de waarde van deze variabele, hoe hoger de overdrijving is. Negatieve waarden op de variabele duiden juist op een grotere onderrapportage. De resultaten laten zien dat kerkelijke respondenten, in deeltijd werkenden en tweeverdieners grotere verschillen laten zien, terwijl respondenten die zeggen vrijwilligerswerk te doen en respondenten van wie de leeftijd onbekend is juist sterker onderrapporteren. Leeftijd, geslacht, opleiding, inkomen en huisbezit vertonen geen significante verbanden met de mate van overdrijving.

## 5. Conclusie en discussie

De vergelijking van giften aan Greenpeace zoals gerapporteerd in een consumentenenquête en geregistreerd door de organisatie laat zien dat er zowel respondenten zijn die giften noemen die niet zijn geregistreerd als respondenten die giften die

**Tabel 3**  
**Regressie van het verschil tussen gerapporteerde en geregistreerde giften**

	B	B
Vrouw	-1,0	-2,0
Ouder dan 65	-2,7	-2,3
Leeftijd onbekend	-10,0(*)	-9,5(*)
Tweeverdiener	7,9(*)	7,4(*)
Parttime werk	11,8*	12,0*
Opleidingsniveau	-0,5	-0,6
Partner in huishouden	7,9	6,3
Inkomen (log)	5,1	1,0
Eigen huis	-4,3	-4,4
Kerkgang	9,9*	10,3*
For-profit dienstensector	5,1	5,9
Gestemd bij laatste verkiezingen		-8,2
Bloeddonor		1,8
Vrijwilliger		-8,0*
Totaal gegeven bedrag (log)		5,7**
Vertrouwen in goededoelenorganisaties		2,3
Constante	-18,9	-23,1
Adj. R <sup>2</sup>	,028	,058

\*\*\* p<.001; \*\* p<.01; \* p<.05; (\*) p<.10.

zijn geregistreerd verzuimen te rapporteren. Sommige respondenten die giften rapporteren overdrijven terwijl andere respondenten juist onderrapporteren. De totale vertekening is nagenoeg nul doordat vergeten en onderrapporteren en het onterecht claimen en overdrijven elkaar nagenoeg geheel opheffen. Dit resultaat is opmerkelijk gezien de resultaten van eerdere studies naar de nauwkeurigheid van gerapporteerde giften en ander sociaalwenselijk gedrag, waarin 'virtually all the error is in the socially desirable direction' (Presser & Traugott, 1992).

Het is moeilijk een definitieve verklaring te geven voor het lage niveau van sociaalwenselijke antwoorden in de huidige analyses. In dit onderzoek gaat het over een andere organisatie en een andere bron van enquêtegegevens, met een andere steekproef dan in eerder onderzoek. De steekproefsamenstelling kan gedeeltelijk een verklaring bieden. De categorieën van respondenten die in de enquête die hier is gebruikt oververtegenwoordigd zijn (vrouwen, lager- en middelbaar opgeleiden, mensen zonder betaald werk) zijn volgens de resultaten van de analyses iets beter in het rapporteren van giften dan de categorieën die ondervertegenwoordigd zijn. De verbanden zijn echter niet groot en evenmin significant. Een andere gedeeltelijke verklaring is dat in voorgaande studies niet altijd de *false negatives* zijn meegeteld. In studies waarin dat wel is gedaan (bijvoorbeeld Burt & Popple, 1998) wordt sociaalwenselijk gedrag echter doorgaans nog steeds overdreven. Een volledige verklaring vormen de steekproefsamenstelling en het meenemen van de *false negatives* dus niet. Hoe dan ook, het lijkt wel veilig om te concluderen dat de uitgebreide *online* vragenlijst het niveau van sociaalwenselijke antwoorden in ieder geval niet heeft vergroot.

Natuurlijk kent ook de opzet van het hier gepresenteerde onderzoek problemen. De respons is erg laag door het gebruik van uitnodigingen per e-mail. Ook is de respons erg selectief, vooral voor wat betreft de arbeidssituatie en het opleidings-

niveau. Wellicht ook hebben respondenten die beter op de hoogte zijn van hun giften aan goededoelenorganisaties de enquête vaker ingevuld. Een ander probleem is dat respondenten 'het afgelopen jaar' geïnterpreteerd kunnen hebben als 'het afgelopen kalenderjaar'. Hierdoor kan de totale nauwkeurigheid onderschat zijn. Het is ook mogelijk dat respondenten sinds de giften zijn gedaan zijn verhuisd, waardoor de nauwkeurigheid eveneens wordt onderschat.

Het resultaat dat de gerapporteerde donaties lager zijn dan geregistreerde donaties onder de respondenten die zowel in de enquête als in de registratiedata voorkomen komt wel overeen met de resultaten van eerder onderzoek (Bekkers & Wiepking, 2010; Burt & Popple, 1998). De overdrijving die nu gevonden is (14%) ligt echter lager dan de overdrijving in eerdere studies (30%). Hoewel onduidelijk is waarom de overdrijving nu kleiner is, is het een geruststellende bevinding. Eveneens in overeenstemming met eerder onderzoek is de bevinding dat kleinere giften vaker worden vergeten dan grotere. Deze bevinding ondersteunt de hypothese dat saillante gebeurtenissen nauwkeuriger worden gerapporteerd.

De bevindingen over kenmerken van respondenten die samenhangen met discrepanties tussen gerapporteerde en geregistreerde giften wijken af van eerdere bevindingen van Bekkers & Wiepking (2010). Nu worden geen significante verbanden gevonden met geslacht, opleiding of inkomen. Wel vinden we een positief verband met kerkgang, terwijl eerder een negatief verband werd gevonden met kerkelijkheid. Het is onduidelijk hoe de resultaten verklaard kunnen worden. Kerkgangers zijn doorgaans vrijgevinger dan onkerkelijken (Bekkers & Wiepking, 2007).

De resultaten bieden steun voor de algemene stelling dat geheugenproblemen en een gebrek aan informatie een belangrijke bron van vertekening zijn in enquêtes. Oudere respondenten gaven minder vaak nauwkeurige antwoorden. Dit kan verklaard worden door afnemend cognitief functioneren. Ook tweeverdieners gaven minder vaak nauwkeurige antwoorden. Deze respondenten zijn waarschijnlijk niet altijd goed op de hoogte van het geefgedrag van de partner.

Een opvallende bevinding is dat de nauwkeurigheid van rapportages over geefgedrag hoger is onder respondenten die ook ander sociaalwenselijk gedrag rapporteren zoals het stemmen bij verkiezingen en het geven van bloed. In enquêtes ligt het percentage respondenten dat zegt gestemd te hebben bij de verkiezingen en het percentage bloed donoren doorgaans hoger dan het populatiegemiddelde. Ook in de huidige gegevens is dat het geval. Vaak worden deze discrepanties geïnterpreteerd als sociaalwenselijk gedrag in enquêtes. Nu blijkt dat de respondenten die zeggen gestemd te hebben en bloed te geven juist de meer nauwkeurige respondenten zijn en minder vaak sociaalwenselijk te antwoorden. Voor respondenten die zeggen vrijwilligerswerk te doen geldt zelfs dat zij de giften onderdrijven. Deze respondenten zijn dus te bescheiden. Het ligt voor de hand dat de oververtegenwoordiging van stemmers en bloeddonors (en waarschijnlijk ook vrijwilligers, hoewel daar geen populatiecijfers voor bekend zijn) onder de NDP-

respondenten veroorzaakt wordt door hun grotere bereidheid aan enquêtes mee te werken (Abraham, Helms & Presser, 2009). Geruststellend is ook het verband tussen nauwkeurigheid en het vertrouwen in goededoelenorganisaties. Het vertrouwen dat deze respondenten zeggen te hebben is geen sociaalwenselijk antwoord. Aanvullende analyses op basis van de geregistreerde giften (hier niet gerapporteerd) laten bovendien zien dat stemmers, bloeddonors, vrijwilligers en respondenten met meer vertrouwen in goededoelenorganisaties ook daadwerkelijk vaker aan Greenpeace geven. Het zijn de betrokken burgers in Nederland die vaker geven, vaker meedoen aan enquêtes en daarin ook nog de meest nauwkeurige antwoorden geven.

## Referenties

Abraham, K.G., Helms, S. and Presser, S. (2009). How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States. *American Journal of Sociology*, 114: 1129–1165.

Bekkers, R., and Wiepking, P. (2007). Generosity and Philanthropy: A Literature Review (October 28, 2007). Available at SSRN: <http://ssrn.com/abstract=1015507>.

Bekkers, R., and Wiepking, P. (2010). 'Accuracy of Self-reports on Donations to Charitable Organizations'. *Quality & Quantity*, 44.

Belli, R.F., Traugott, M.W., Young, M., and McGonagle, K.A. (1999). Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring. *Public Opinion Quarterly*, 63: 90–108.

Bernstein, R., Chadha, A., and R. Montjoy (2001). Overreporting Voting: Why It Happens and Why It Matters. *Public Opinion Quarterly*, 65: 22–44.

Burt, C.D.B., and Poppo, J.S. (1998). Memorial Distortions in Donation Data. *Journal of Social Psychology*, 138: 724–733.

Cahalan, D. (1968). Correlates of Respondent Accuracy in the Denver Validity Survey. *Public Opinion Quarterly*, 32: 607–621.

CBF (2009). *Financiële resultaten van goededoelenorganisaties in Nederland*. Verslag Fondsenwerving 2008. Amsterdam: CBF.

Crowne, D.P., and Marlowe, D. (1960). A New Scale of Social Desirability Independent of Psychopathology. *Journal of Consulting Psychology*, 24: 349–354.

Crowne, D.P., and Marlowe, D. (1964). *The Approval Motive: Studies in Evaluative Dependence*. New York: John Wiley & Sons.

Hahn, R.A., Eaker, E., and Rolka, H. (1997). Reliability of reported age at menopause. *American Journal of Epidemiology*, 146: 771–775.

Hardt, J., and Rutter, M. (2004). Validity of adult retrospective reports of adverse childhood experiences: review of the evidence. *Journal of Child Psychology and Psychiatry*, 45 (2): 260–273.

Harrell, A. (1997). The validity of self-reported drug use data: the accuracy of responses on confidential self-administered answered sheets. *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*, ed. Harrison, L., and Hughes, A., Rockville: U.S. Department of Health and Human Services.

Johnson, T.P., O'Rourke, D.P., Burris, J.E., and Warnecke, R.B. (2005). An investigation of the effects of social desirability on the validity of self-reports of cancer screening behaviors. *Medical Care*, 43 (6): 565–573.

Katosh, J.P., and Traugott, M.W. (1981). The Consequences of Validated and Self-Reported Voting Measures. *Public Opinion Quarterly*, 45: 519–535.

Maxfield, M.G., Weiler, B.L., and Widom, C.S. (2000). Comparing self-reports and official records of arrests. *Journal of Quantitative Criminology*, 16: 87–110.

Midanik, L. (1982). The Validity of Self-Reported Alcohol Consumption and Alcohol Problems: A Literature Review. *British Journal of Addiction*, 77: 357–382.

Mills, J.F., and Kroner, D.G. (2005). An Investigation Into the Relationship Between Socially Desirable Responding and Offender Self-Report. *Psychological Services*, 2: 70–80.

Parry, H.J., and Crossley, H.M. (1950). Validity of Response to Survey Questions. *Public Opinion Quarterly*, 14, 61–80.

Phillips, D.L., and K.J. Clancy (1972). Some Effects of “Social Desirability” in Survey Studies. *American Journal of Sociology*, 77: 921–940.

Presser, S. (1984). Is Inaccuracy on Factual Survey Items Item-Specific or Respondent-Specific? *Public Opinion Quarterly*, 48: 344–355.

Presser, S., and Stinson, L. (1998). Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance. *American Sociological Review*, 63: 137–145.

- Presser, S., and Traugott, M. (1992). Little White Lies and Social Science Models: Correlated Response Errors in a Panel Study of Voting. *Public Opinion Quarterly*, 56: 77–86.
- Schuyt, T.N.M., Gouwenberg, B.M., and Bekkers, R. (2009). *Geven in Nederland 2009: Giften, Sponsoring, Legaten en Vrijwilligerswerk*. Amsterdam: Reed Business.
- Silver, B.D., Anderson, B.A., Abramson, P.R. (1986). Who Overreports Voting? *American Political Science Review*, 80: 613–624.
- Stocké, V. (2007a). Response Privacy and Elapsed Time Since Election Day as Determinants for Vote Overreporting. *International Journal of Public Opinion Research*, 19: 237–246.
- Stocké, V. (2007b). Determinants and Consequences of Survey Respondents' Social Desirability Beliefs about Racial Attitudes. *Methodology*, 3: 125–138.
- Stoop, I.A.L. (2005). *The Hunt for the Last Respondent*. The Hague: SCP.
- Sudman, S., and Bradburn, N.M. (1973). Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association*, 73: 805–815.
- Thornberry, T.P., and Krohn, M.D. (2000). *The Self-Report Method for Measuring Delinquency and Crime*, in *CJ 2000: Innovations in Measurement and Analysis*, ed.
- Duffee, D., Crutchfield, R.D., Mastrofski, S., Mazerolle, L., McDowall, D., and Ostrom, B., Washington, DC: National Institute of Justice.
- Tourangeau, R., Jobe, J.B., Pratt, W.F., and Rasinski, K. (1997). Design and Results of the Women's Health Study. In *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*, ed. Harrison, L., and Hughes, A., Rockville: U.S. Department of Health and Human Services.
- Tourangeau, R. Rips, L.J., and Rasinski, K.A. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Tourangeau, R., and Smith, T. (1996). Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly*, 60: 275–304.
- Traugott, M.W., and Katosh, J.P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly*, 43: 359–377.

Udry, J.R., Gaughan, M., Schwingl, P.J., and Van den Berg, B. (1996). A Medical Record Linkage Analysis of Abortion Underreporting. *Family Planning Perspectives*, 28: 228–231.

Wilhelm, M.O., and Bekkers, R. (2010). Helping Behavior, Dispositional Empathic Concern, and the Principle of Care. *Social Psychology Quarterly*, 73: 1–22.

WWAV (2009). *Onderzoeksrapportage Het Nederlandse Donateurspanel*, December 2009. Woerden: WWAV/NETQ/WDM Nederland.

## *Noten*

- 
- <sup>1)</sup> René Bekkers, Center for Philanthropic Studies, Faculty of Social Sciences, VU University Amsterdam. De Boelelaan 1081, 1081 HV Amsterdam, the Netherlands. r.bekkers@vu.nl  
Dit artikel zou niet mogelijk zijn geweest zonder de inzet van Jan van Berkel, Baukje Stam, Dana Henny en Remco Meyaard van WWAV en door Kees Zitman en Willem van den Berg van Greenpeace. Dit artikel is een korte versie van het paper 'Accuracy of Self-Reported Donations to Charitable Organizations', gepresenteerd op het congres 'Registers in sociaalwetenschappelijk onderzoek: mogelijkheden en valkuilen', 24 juni, Leiden.



# Een buitenkansje!

## *Discussie over Bekkers, R. (VU), Nauwkeurigheid van metingen in registers en enquêtes*

Jelke Bethlehem (UvA/CBS)

### 1. *Inleiding*

Als er behoefte bestaat aan informatie over een bepaald onderwerp, dan kan worden besloten om daarvoor een survey uit te voeren. Surveys zijn echter tijdrovend en kostbaar. Daarom verdient het aanbeveling om eerst eens uit te zoeken of niet mogelijk is om uit al bestaande bronnen gegevens te verzamelen. Registers en administraties kunnen hiervoor in aanmerking komen. Een voordeel van dit soort bronnen is dat er geen kosten meer hoeven te worden gemaakt voor het verzamelen van de gegevens. Een nadeel is dat de onderzoeker geen controle heeft gehad over de wijze waarop de gegevens zijn verzameld. Dus is het ook niet altijd duidelijk wat de kwaliteit van de gegevens is. Het is daarom zeer interessant dat Bekkers in staat is geweest om survey-gegevens te vergelijken met gegevens uit een administratie. Dit soort buitenkansjes krijg je niet vaak.

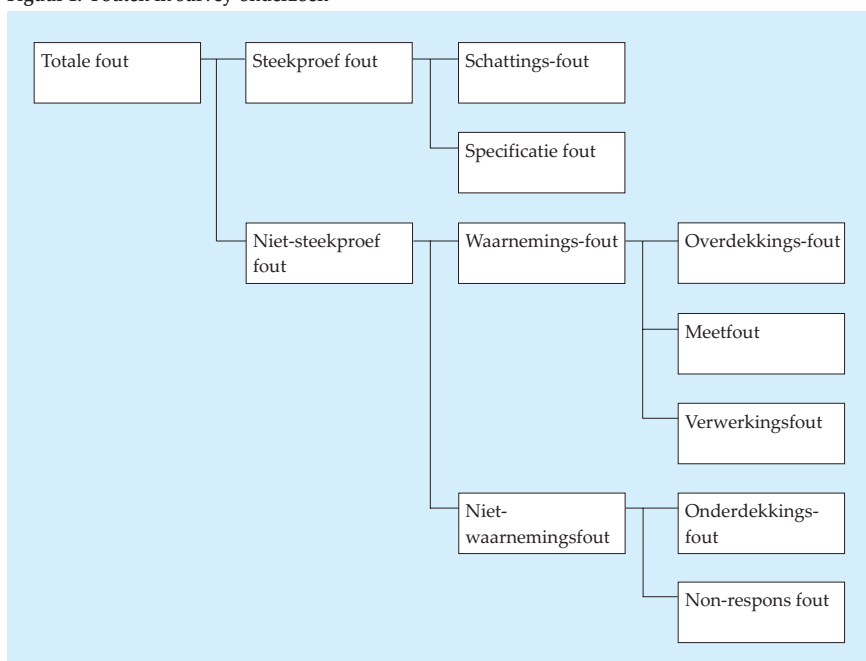
### 2. *Over fouten in gegevens*

In een survey heeft Bekkers vragen kunnen stellen over donaties aan specifieke organisaties voor een goed doel: of men had gedoneerd en hoeveel. De uitkomsten van de survey kon hij vergelijken met de financiële administratie van de organisatie.

Zowel met gegevens uit surveys als gegevens uit registers en administraties kan van alles mis zijn. Bethlehem (2009) geeft een schematisch overzicht van mogelijke fouten in surveys. Dit overzicht is gereproduceerd in figuur 1.

Alle problemen die vallen onder steekproeffouten, kunnen zich alleen voordoen bij surveys. De niet-steekproeffouten komen voor in surveys en registers en administraties. Het onderzoek van Bekkers beperkt zich tot maar één aspect, en dat is meetfouten. Dat zijn fouten die ontstaan doordat niet de juiste waarde van de te onderzoeken variabele wordt vastgelegd. Afwijkingen kunnen allerlei oorzaken hebben. In het onderzoek van Bekkers komen twee soorten meetfouten voor: geheugenefecten en sociaal wenselijke antwoorden.

Figuur 1. Fouten in survey-onderzoek



### 3. *De steekproef*

De steekproef voor de survey over het donatiegedrag is via een aantal stappen ontstaan. Uitgangspunt was de Grote Consumentenenquête. Deze enquête wordt uitgezet bij alle huishoudens. Respondenten kunnen de vragen beantwoorden via een papieren vragenlijst of via een vragenlijst op het web. De respons van deze enquête is slechts 200 000 huishoudens. Uit deze responderende huishoudens is een steekproef getrokken voor een donatiepanel van 60 000 huishoudens. De respons daarvan blijkt 2 300 huishoudens. Al met al kunnen vraagtekens worden gezet bij de representativiteit van deze survey. Weliswaar ging het niet daarom in het onderhavige onderzoek, maar je kunt je toch afvragen of de gevonden patronen ook zouden zijn aangetroffen bij de non-respondenten. De vraag is daarmee in hoeverre de conclusies uit het onderzoek van Bekkers kunnen worden gegeneraliseerd.

In zijn analyse gaat Bekker ervan uit dat eventuele verschillen tussen uitkomsten in de survey en in de administratie het gevolg zijn van fouten in de survey. De administratie wordt dus als benchmark gebruikt. Dit impliceert dat de administratie geen fouten bevat. Dat is in dit geval vermoedelijk wel redelijk omdat het gaat om de financiële administratie van een goede doelen organisatie, maar in het algemeen is het niet vanzelfsprekend dat registers en administraties foutloos zijn.

#### 4. *Geheugeneffecten*

Bekkers vindt vooral geheugeneffecten in de survey. Dat is een probleem dat zich vaak voordoet bij vragen die een beroep doen op de herinnering. Mensen vergeten gebeurtenissen (vooral als ze minder belangrijk zijn en regelmatig voorkomen) of plaatsen gebeurtenissen verkeerd in de tijd (telescoping). Ook in de donatiesurvey worden veel (kleine) donaties vergeten.

Bethlehem (2009) geeft een voorbeeld van ander onderzoek waarin geheugeneffecten een grote rol spelen. Hij beschrijft CBS-onderzoek van Sikkel (1983) waarin aan mensen werd gevraagd hoe vaak ze in de afgelopen 3 maanden naar de huisarts zijn geweest. Uit het onderzoek bleek dat gemiddeld een kwart van de doktersbezoeken werd vergeten.

De geheugenproblemen nemen toe naarmate de periode langer is waarover moet worden gerapporteerd. In de survey van Bekkers moesten de respondenten rapporteren over het afgelopen jaar. Die periode is vermoedelijk te lang. Het zou interessant kunnen zijn om te experimenteren met een vraagstelling voor een kortere periode.

#### 5. *Sociaal-wenselijke antwoorden*

Bekkers vindt ook sociaal-wenselijke antwoorden. In de survey worden donaties gerapporteerd die niet zijn terug te vinden in de administratie van de organisatie. De analyse van dit verschijnsel duidt erop dat dit zich vooral voordoet bij religieuze mensen.

De literatuur op het gebied van survey-onderzoek lijkt erop te duiden dat sociaal-wenselijke antwoorden vooral voorkomen bij surveys waarbij interviewers aanwezig zijn bij het invullen van de vragenlijst. Bij schriftelijke surveys en web-surveys zouden de effecten van sociaal-wenselijke antwoorden minder moeten zijn. Het zou interessant kunnen zijn om uit te zoeken of het effect van sociaal-wenselijke antwoorden inderdaad toeneemt als interviewers worden ingezet bij het verzamelen van de gegevens over donaties.

Overigens hoeven interviewers niet alleen maar een negatief effect te hebben op de uitkomsten van survey-onderzoek. Zo blijkt het inzetten van interviewers meestal tot een veel hogere respons te leiden.

#### 6. *Ten slotte*

De afwijkingen die door geheugeneffecten ontstaan lijken gemiddeld weg te vallen tegen de afwijkingen veroorzaakt door sociaal-wenselijke antwoorden. Dat lijkt mooi, want er ontstaat zo geen vertekeningen in schattingen op het niveau van de complete respons. Dit geeft natuurlijk geen garantie dat er ook geen effecten in al-

lerlei deelgroepen zullen kunnen optreden. Zo zijn de effecten van sociaal-wenselijke antwoorden vooral bij religieuze mensen sterker dan bij andere mensen. Dus als deze groep apart wordt geanalyseerd, dan zouden best wel eens vertekeningen kunnen optreden.

Het onderzoek van Bekkers toont aan dat het belangrijk is om gegevens in surveys en administraties met elkaar te vergelijken. Dit kan belangrijke inzichten opleveren in de kwaliteit van de verzamelde gegevens. Zulke buitenkansjes doen zich niet vaak voor. Dus als ze zich voordoen, dan moet de gelegenheid zeker te baat worden genomen.

## *Referenties*

Bethlehem, J. (2009). *Applied Survey Methods, A Statistical Perspective*. Hoboken, Nj, USA: John Wiley & Sons.

Sikkel, D. (1983), Geheugeneffecten bij het rapporteren van huisartsencontacten. *Statistisch Magazine* 3, nr. 4, *Netherlands Central Bureau of Statistics*, pp.61–64.

# *Koppeling van registers: Onmisbaar voor een beter inzicht in de volksgezondheid*

Anton Kunst (AMC)

## *1. Inleiding en doelstelling*

Het CBS heeft zich de afgelopen jaren grote moeite getroost om verschillende registers te koppelen in het kader van de Sociaal Statistische Bestand (SSB) en in het verlengde daarvan het Gezondheid Statistisch Bestand (GSB). Veel werk is verricht om methodische problemen op te lossen zoals het ontbreken van persoonsidentificatienummers, inconsistenties tussen gegevens uit verschillende registers, missende waarden in vooral administratieve bestanden, en discrepanties in definities of classificaties.

Ik zal in mijn bijdrage niet ingaan op deze problemen, maar uitgaan van het gegeven dat dankzij de inspanningen van het CBS en haar samenwerkingpartners een prachtig systeem is ontwikkeld van registers die op individueel niveau aan elkaar gekoppeld kunnen worden. Dit systeem biedt niet alleen de mogelijkheid betere statistieken uit te brengen, maar biedt bovendien voor onderzoekers nieuwe mogelijkheden voor grensverleggend onderzoek. Toegang tot gekoppelde registerbestanden is mede te danken aan maatregelen die door en op het CBS zijn getroffen om zorgen over bescherming van privacy weg te nemen. Nu kan onderzoek worden uitgevoerd waar onderzoekers 10 tot 15 jaar geleden alleen maar van konden dromen. Ik ben een van die mensen die de nieuwe mogelijkheden met grote dankbaarheid heeft aangegrepen.

Mijn onderzoeksterrein is dat van de volksgezondheid. Dit onderzoek beoogt de gezondheidssituatie van de Nederlandse bevolking in kaart te brengen, met bijzondere aandacht voor ontwikkelingen over de tijd en ongelijkheden tussen bevolkingsgroepen. Op basis van mede dit onderzoek worden prioriteiten en strategieën voor zorgbeleid en collectieve preventie geformuleerd. Het onderzoek op dit terrein evalueert ook de effecten van nieuwe beleidsmaatregelen op de volksgezondheid, in het verleden of in de toekomst (Mackenbach en van der Maas, 2008). Binnen dit onderzoekerrein speelt onderzoek op basis van gekoppelde registers een steeds grotere rol. Verwacht wordt dat die rol in de toekomst alleen maar zal toenemen.

Ik wil vanuit mijn ervaring op dit onderzoeksterrein reflecteren op de mogelijkheden die koppeling van registers de afgelopen jaren hebben geboden, en de nieuwe mogelijkheden die in het vooruitzicht liggen. In sectie 2 zal ik in vogelvlucht een

overzicht geven van het onderzoek dat in de afgelopen jaren is uitgevoerd op basis van koppeling van registers. In sectie 3 zal ik uiteenzetten in welke opzichten dat onderzoek heeft kunnen profiteren van de mogelijkheden van gekoppelde registergegevens. In sectie 4 zal ik het voorbeeld geven van een specifiek onderzoeksproject, om aan te geven dat registergegevens niet alleen mogelijkheden bieden, maar ook hun beperkingen hebben. In sectie 5 formuleer ik aanbevelingen voor het toekomstige gebruik van gekoppelde registergegevens door onderzoekers.

## 2. *Ontwikkeling in onderzoek naar sterfte in Nederland*

Ik zal mij in deze sectie richten op onderzoek naar sterfte in Nederland. Uiteraard is sterfte slechts een van de vele uitkomstmaten binnen de volksgezondheid, naast kwaliteit van leven, ziekten, risicofactoren, toegankelijkheid van zorgvoorzieningen, en kwaliteit van zorg. Het voorbeeld van de sterfte laat echter goed zien hoe in de afgelopen jaren het Nederlands onderzoek vooruitgang heeft geboekt dankzij het beschikbaar komen van gegevens door koppeling van registers.

Het overzicht begint in sectie 2.1 met sterfteonderzoek van vóór 1995, toen koppeling van landelijke registers nog niet mogelijk was en het sterfte-onderzoek was gebaseerd op de sterfteregistratie en de doodsoorzaakstatistiek. Secties 2.2 en 2.3 bespreken hoe dit onderzoek later is verrijkt door koppeling van de sterftegegevens aan de Gemeentelijke Basis Administratie (GBA) respectievelijk aan andere registers of enquêtes zoals de Enquête Beroepsbevolking (EBB).

### 2.1. *Onderzoek vóór 1995*

De sterftestatistiek heeft van oudsher een centrale plaats ingenomen in onderzoek naar de volksgezondheid. Gegevens over sterfte naar leeftijd en geslacht zijn sinds het midden van 19<sup>de</sup> eeuw beschikbaar. Op basis van deze gegevens wordt bijvoorbeeld jaarlijks de levensverwachting bij geboorte gepubliceerd. Deze levensverwachting is de afgelopen 150 jaar gestaag toegenomen. Ook eind 20<sup>ste</sup> eeuw ging die toename door (van der Lucht en Polder, 2010). Tussen 1950 en 2005 is per 7 kalenderjaren de levensverwachting met 1 jaar toegenomen. Deze winst in levensverwachting is vaak aangehaald als een kernindicator waaraan de vooruitgang in de volksgezondheid kon worden afgelezen. Nog altijd wordt binnen het vakgebied van de public health verwezen naar de successen die aan deze vooruitgang ten grondslag lagen, zoals de succesvolle strijd tegen de infectieziekten in het begin 20<sup>ste</sup> eeuw, en de daling van meer dan 75 procent in sterfte aan hartinfarct sinds 1970 (Mackenbach en van der Maas, 2008).

Het belang van sterfte-onderzoek wordt ook onderstreept door het feit dat de levensverwachting in Nederland ondanks de toename in de afgelopen jaren toch is achtergebleven bij de rest van West Europa (Janssen et al, 2004). Nu verkeert de Nederlandse levensverwachting in de Europese middenmoot. Dit gegeven drukt

een flinke stempel op het beeld van de Nederlandse volksgezondheid, zoals dat bijvoorbeeld elke 4 jaar door het RIVM wordt uitgedragen in de Volksgezondheid Toekomst Verkenning. Die positie in de middenmoot heeft in de Nederlandse gezondheidssector de vraag opgeroepen wat wij verkeerd doen, en wat wij kunnen leren van het buitenland (van der Lucht en Polder, 2010).

In sterfteonderzoek van vóór 1995 is vooral gebruik gemaakt van de mogelijkheid om gegevens op regionaal niveau te analyseren. In bijvoorbeeld een analyse van sterfte aan verkeersongevallen is gebleken dat deze sterfte vooral hoog was in landelijke gebieden. Een vergelijking (niet: koppeling) met gegevens uit ongevalregistraties wees uit dat de landelijke gebieden zich niet onderscheiden door een hogere incidentie van verkeersongevallen, maar wel door een hogere overlijdenskans in geval van een ongeval. Deze hogere *case-fatality* was mede te wijten aan de fysieke infrastructuur van landelijke gebieden, met bijvoorbeeld veel 80 km wegen en gelijkwaardige kruisingen (van Beeck et al, 1991).

Vóór 1995 is tevens een start gemaakt met de analyse van sterfte op het niveau van gemeenten en buurten. Dit geografisch onderscheid is onder meer benut in het Vergelijkend Buurt Onderzoek (VBO) dat eind jaren 1970 in Amsterdam is uitgevoerd. Een kernbevinding uit dit onderzoek was dat mensen in de armste wijken een 5 jaar kortere levensverwachting hadden dan mensen in de rijkste wijken (Lau-IJzerman et al, 1981). Dit gegeven is tot in de jaren 1990 met grote regelmaat in publieke discussies aangehaald om te benadrukken dat sociaal-economische verschillen in sterfte in Nederland aanzienlijk waren en dat daar iets aan moest worden gedaan. Dankzij het onderscheid naar buurt kon het VBO dus de weg bereiden tot een beleidsthema, dat van ongelijkheid in gezondheid, dat nu niet meer weg is te denken uit onderzoek en beleid op het terrein van de volksgezondheid in Nederland (Mackenbach en Stronks, 2004).

Kortom, vóór 1995 werd sterfte in Nederland vooral onderzocht in relatie tot leeftijd, geslacht en geografische variabelen. In sommige studies werd ook gekeken naar verschillen in relatie tot burgerlijke staat of nationaliteit (van Poppel en Joung, 2001). Veel inzichten werden bovendien ontleend aan analyses van specifieke doodsoorzaken. Midden jaren 1990 werden ook pogingen ondernomen om informatie over “secundaire” oorzaken te benutten (Kunst et al, 2003). Maar daarmee waren de mogelijkheden van de traditionele sterftestatistiek zo goed als uitgeput.

## 2.2. *Onderzoek op basis van koppeling met GBA*

In 1995 werd de GBA landelijke geautomatiseerd en was koppeling met de sterfteregistratie mogelijk geworden. Dit bood nieuwe mogelijkheden voor sterfteonderzoek, vooral op het terrein van sterfteverschillen. Bijvoorbeeld het onderzoek naar buurtverschillen in sterfte, zoals uitgevoerd binnen Amsterdam, kon eind jaren 1990 voor het eerst ook op landelijk niveau worden uitgevoerd. De eerste van deze studies toonde bijvoorbeeld aan dat sterfteverschillen tussen arme en rijken buurten overal in Nederland bestonden. Bovendien bleek dat deze

verschillen in de zuidelijke provincies groter waren dan in andere landsdelen (Bos et al, 2002).

Een groot winstpunt was dat vanaf 1995 sterfte kon worden onderzocht in relatie tot land van herkomst. Sinds ongeveer 2000 zijn een groot aantal studies verschenen naar sterfte onder Nederlanders van buitenlandse herkomst, in het bijzonder van niet-westerse herkomst. Veel van deze studies richtten zich op de vier klassieke migrantengroepen (Surinamers, Marokkanen, Turken en Antillianen/Arubanen). Tegenwoordig is voor deze groepen het niveau van sterfte gedetailleerd in kaart gebracht (Garssen et al, 2003). Maar ook het niveau en het patroon van sterfte van mensen van Indonesische afkomst, of van mensen uit Afrika en West Azië, is in enkele studies onderzocht (Bos, 2005).

De mogelijkheden die ontstonden door koppeling van de sterfteregistratie met de GBA werden het meest uitgebuit in een studie naar sterfte aan verschillende kankersoorten (Stirbu et al, 2006). Op basis van de zogenaamde “convergentie hypothese” werd onderzocht of het niveau van kankersterfte onder allochtonen samenhang met hun generatie (eerste of tweede), hun leeftijd van migratie naar Nederland, en hun verblijfsduur in Nederland. Conform de verwachting bleek het niveau van kankersterfte dicht bij het Nederlandse niveau onder de tweede generatie migranten, of onder eerste generatie migranten die lang geleden en/of op jonge leeftijd naar Nederland waren gekomen. Dit zijn immers de migranten die een relatief groot deel van hun leven waren blootgesteld aan de leefomstandigheden en de leefgewoonten in Nederland. Maar convergentie bleek een langzaam proces te zijn: ook tweede generatie migranten hadden nog een duidelijk lagere kans op sterfte aan kanker dan bewoners van Nederlandse afkomst.

Dankzij de landelijke koppeling met de GBA waren ook andere stratificaties mogelijk geworden. Het CBS heeft bijvoorbeeld onderzocht welke demografische factoren ten grondslag lagen aan de, vanuit Europees perspectief, vrij hoge perinatale sterfte in Nederland (Garssen en van der Meulen, 2004). Dit kon worden onderzocht door perinatale sterfte te onderzoeken in relatie tot kenmerken van de moeder zoals haar leeftijd bij geboorte. De hogere sterfte in Nederland bleek samen te hangen met het feit dat moeders in Nederland relatief oud waren bij de geboorte van hun eerste kind, maar vooral ook met een groot aandeel geboorten onder niet-westerse migranten. Deze analyse gaf daarmee een belangrijke bijdrage aan discussies over vraag waardoor perinatale sterfte in Nederland relatief hoog is, onder meer door duidelijk te maken dat er meerdere factoren in het spel waren dan alleen problemen in de gezondheidszorg.

Een ander voorbeeld betreft de toename van sterfte in de Nederlandse bevolking tijdens hittegolven (Garssen et al, 2005). Hittegerelateerde sterfte was in het begin van de jaren 1990 al waargenomen in Nederlands onderzoek op basis van de sterftestatistiek. Dit onderzoek had aannemelijk gemaakt dat een hoge buitentemperatuur een direct effect had op overlijdenskansen van ouderen. Onzeker was echter welke groepen in de Nederlandse samenleving het meeste risico liepen tijdens hit-



tegolven. Uit Frans onderzoek kwam naar voren dat alleenstaande ouderen een risicogroep vormden. In later CBS onderzoek bleek echter dat in Nederland onder bewoners van instellingen als verzorgings- en verpleeghuizen de sterfte tijdens hittegolven veel meer toenam dan onder zelfstandig wonende ouderen. Met deze bevinding werd een beeld uit Frans onderzoek gecorrigeerd. Het liet zien dat voor de preventie van hittegerelateerde sterfte men zich vooral zal moeten richten op de bescherming van bewoners van instellingen.

### 2.3. *Onderzoek op basis van koppeling aan overige registers*

In recente jaren zijn verschillende studies verschenen waarin sterftestatistiek is gekoppeld met andere registers dan alleen de GBA. In deze sectie geef ik vier voorbeelden van de waarde van dat onderzoek. Met behulp van fiscale bronnen is het mogelijk geworden sterftetekansen van mensen te onderzoeken in relatie tot het recent genoten inkomen, gemeten op persoonsniveau of huishoudenniveau. Een eerste poging in deze richting werden al in het midden van de jaren 1990 ondernomen door het CBS (Kardaun en Glerum, 1995). Een aantal problemen met koppeling van registers werden toen geëvalueerd. Op basis daarvan werd geconcludeerd dat het nog niet mogelijk was voor Nederland als geheel schattingen te maken van arm-rijk verschillen in overlevingskansen. Daarmee bleef Nederland in de jaren 1990 een van de weinig West Europese landen waarvoor het niet mogelijk was schattingen van sociaaleconomische verschillen in sterfte te maken op basis van integrale of landelijk representatieve gegevens (Mackenbach et al, 2008).

In recente jaren zijn de mogelijkheden voor koppeling met fiscale bronnen echter uitgebreid. Sinds 2009 verschenen schattingen van het CBS van de omvang van sterfteverschillen tussen inkomensgroepen, waarbij gebruik is gemaakt van integrale inkomensgegevens (Knoops en van den Brakel, 2010). Deze schattingen werden uitgedrukt in termen van levensverwachting bij geboorte. Het verschil tussen de hoogste en laagste inkomensgroepen in levensverwachting bij geboorte bedroeg ongeveer 5 jaar, zowel onder mannen als vrouwen. Deze schattingen hebben eind 2008 veel aandacht gekregen in de media en hebben in 2009 geholpen het onderwerp van sociaaleconomische gezondheidsverschillen op de politieke agenda te houden.

Tot voor kort was het een belangrijke vraag hoe groot de sterfteverschillen zijn in relatie tot opleiding. Deze sociaaleconomische indicator is het meest gebruikt in Nederlands onderzoek naar sociaaleconomische verschillen in gezondheid, vooral in epidemiologische studies (Kunst, 2007). Een nadeel van deze studies was de vaak kleine steekproefomvang en de beperking tot lokale populaties zoals de wijk Ommoord in Rotterdam (van Hertem et al, 2002). Het CBS heeft echter in 2008 de sterfteregistratie gekoppeld aan gegevens uit de EBB over de voltooide opleiding, voor een landelijk representatieve steekproef van een groot aantal personen (Stam et al, 2008). Tussen de hoogste en laagste opleidingsgroepen bedroeg het verschil in levensverwachting bij geboorte 6,9 jaar onder mannen en 5,7 jaar onder vrou-

wen. Opvallend was dat dit verschil aanzienlijk groter was dan eerdere schattingen op basis van epidemiologische onderzoeken. Een nadere evaluatie van die discrepantie wees uit dat eerdere studies de omvang van opleidingsverschillen in sterfte hadden onderschat (Kunst en Droomers, 2009). De ware omvang was pas met de nieuwe CBS gegevens aan het licht gekomen.

Naast fiscale bronnen en de EBB is de sterfteregistratie recent ook gekoppeld aan POLS enquêtes. In bijvoorbeeld lopend onderzoek aan het Erasmus MC zijn respondenten uit de 1997–99 POLS enquêtes voor sterfte gevolgd tot het jaar 2007 (Klijs et al, 2010). Dit bood de mogelijkheid om sterftekansen te onderzoeken in relatie tot de verschillende in POLS gemeten factoren. Voor preventie is in het bijzonder van belang de effecten te meten van risicofactoren zoals roken en overgewicht. In de Rotterdamse studie bleek dat mensen met ernstig overgewicht een hogere kans hadden op overlijden. Hun levensverwachting op leeftijd 55 jaar was ongeveer 1 jaar lager dan van mensen met normaal gewicht. Maar mensen met een matig overgewicht (een body mass index tussen 25 en 30) hadden lagere in plaats van hogere sterfte. Hun levensverwachting lag ongeveer een half jaar boven die van mensen met normaal gewicht. Dit onderzoek bevestigde daarmee recente bevindingen uit buitenlands onderzoek, dat matig overgewicht tegenwoordig geen negatieve effecten meer heeft op de levensverwachting van ouderen (Bonneux en Reuser, 2007). Die effecten moeten tegenwoordig vooral worden gezocht in een grotere ziektelast tijdens het leven van mensen met overgewicht.

Het laatste voorbeeld betreft de koppeling met registraties uit de zorg. In de afgelopen 10 jaar heeft het CBS veel geïnvesteerd in koppelingen met de Landelijke Medische Registratie (LMR). Deze registratie bevat gedetailleerde informatie over opnamen in vrijwel alle ziekenhuizen en psychiatrische instellingen in Nederland. Koppeling van de LMR aan de GBA biedt de mogelijkheid om ontslagen patiënten over de tijd te volgen, onder meer wat betreft hun kansen op heropname voor dezelfde diagnose. Ook kunnen ontslagen patiënten worden gevolgd voor sterfte.

Een recent onderzoek heeft zich bijvoorbeeld gericht op patiënten die zijn opgenomen als gevolg van hart- en vaatziekten (Stirbu en Mackenbach, 2009). Deze patiënten zijn gevolgd voor sterfte in een periode tot 1 jaar na de eerste opname. De kans op overlijden bleek fors hoger te zijn onder lagere inkomensgroepen. Onder patiënten van niet-westerse herkomst bleek, anders dan verwacht, de overlijdenskans niet veel hoger te zijn dan onder autochtone patiënten. Een belangrijke vraag bij deze resultaten is of arm-rijk verschillen in overleving na een eerste opname toegeschreven kunnen worden aan verschillen in de kwaliteit van zorg die in het ziekenhuis is ontvangen. Ongelijkheid in de zorgkwaliteit en zorguitkomsten kan niet uitgesloten worden. De studie vond hiervoor echter geen sterke aanwijzingen. Bijvoorbeeld het gebruik van chirurgische cardiovasculaire verrichtingen verschilde niet tussen arme en rijke patiënten.

### 3. *Evaluatie van winstpunten*

De voorbeelden uit sectie 2 laten zien dat het onderzoek naar sterfte in Nederland in de afgelopen jaren zich in snel tempo heeft verbreed en verdiept. Koppeling van de sterfteregistratie aan allerlei registers heeft nieuwe mogelijkheden voor analyse geschapen, mogelijkheden waarvoor men vóór 1995 alleen maar van kon dromen. De voorgaande sectie gaf hiervan enkele voorbeelden. In deze sectie willen wij op basis van die voorbeelden evalueren op welke punten de mogelijkheden voor analyse zijn toegenomen. Wat was, vanuit een methodologisch perspectief, de winst van koppeling van registers? Hieronder onderscheiden wij vier winstpunten.

#### 3.1. *Differentiëren naar persoonskenmerken*

Door koppeling aan registerbestanden met informatie over persoonkenmerken kan nu voor Nederland als geheel worden onderzocht in hoeverre sterfteskansen verschillen tussen bevolkingsgroepen. Vóór 1995 kon onderscheid worden gemaakt naar leeftijd, geslacht, burgerlijke staat, nationaliteit en geografische kenmerken zoals gemeente en buurt. Dankzij koppeling aan de GBA kunnen nu ook land van herkomst, migratiegeschiedenis en kenmerken van het huishouden worden onderzocht in relatie tot sterfte. En sinds enkele jaren kunnen ook sociaaleconomische verschillen in sterfte in kaart worden gebracht, dankzij koppeling met fiscale bronnen en de EBB.

Binnen de bestaande registraties bestaan veel mogelijkheden voor verder onderzoek naar sociale stratificatie en gezondheid (Kunst et al, 2005). Sociaaleconomische verschillen in gezondheid kunnen bijvoorbeeld in meer detail worden beschreven door de mogelijkheden van verschillende registers verder te benutten. Mogelijkheden worden geboden door (a) de EBB en het POLS voor de meting van de arbeidsmarktpositie en beroepsklasse voor grote steekproeven van de werkende bevolking, (b) gegevens uit woningregisters voor het vaststellen van het bezit en de WOZ waarde van woningen waarin mensen wonen en (c) fiscale bronnen voor de meting van het vermogen van huishoudens, inclusief schulden, en in samenhang met het inkomen.

Een lopend CBS onderzoek heeft bijvoorbeeld als doel fiscale bronnen te gebruiken voor een betere identificatie van bevolkingsgroepen waar sociaaleconomische problemen en gezondheidsproblemen accumuleren. Een centrale hypothese in dit onderzoek is dat gezondheidsproblemen vooral veel voorkomen onder mensen die te maken hebben met zowel een laag inkomen als weinig bezittingen. Eerste analyses, op basis van gegevens uit het POLS, bevestigen deze verwachting (van Wingen et al, 2010).

#### 3.2. *Ontleden van gefaseerde processen*

Sterfte is de uitkomst van een proces dat in fasen of stappen kan worden uiteen gefafeld. Ook andere gezondheidsuitkomsten kunnen worden ontleed in stappen die

daar noodzakelijkerwijs er aan voorafgaan. “Decompositie” van zulke processen in opeenvolgende stappen maakt het mogelijk om de uitkomst beter te begrijpen. In bijvoorbeeld onderzoek naar ontwikkelingen in sterfte aan hart- en vaatziekten kan een onderscheid worden gemaakt tussen ontwikkelingen in de klinische incidentie van hartziekten, en ontwikkelingen in de overleving na eerste opname.

Op dezelfde wijze kunnen sociale verschillen in sterfte aan een bepaalde ziekte worden onderzocht als het product van (a) sociale verschillen in incidentie van die ziekte en (b) sociale verschillen in prognose van die ziekte (Stirbu en Mackenbach, 2008). Sociale verschillen in incidentie verwijzen naar de rol van risicofactoren zoals roken en overgewicht. Een belangrijk rol van leefstijlfactoren wijst op mogelijkheden zijn voor preventief beleid. Sociale verschillen in prognose verwijzen naar een mogelijke rol van ongelijke behandeling van hartpatiënten. Als de behandeling verschillende uitkomsten heeft voor rijk en arm, ligt hier een belangrijk aandachtsveld voor zorgbeleid. De rol van gezondheidszorg in het ontstaan van sociaaleconomische verschillen in sterfte is echter nog grotendeels in nevelen gehuld. Koppeling van de GBA, LMR en andere registers zal op dit punt ongetwijfeld veel aan het licht brengen, ook al zal aanvullend diepte-onderzoek nodig zijn voor een volledig antwoord.

Met behulp van medische registers bestaan vele mogelijkheden voor het verder beschrijven van gezondheid in termen van gefaseerde ziekteprocessen. Deze mogelijkheden worden geboden door bijvoorbeeld koppeling van gegevens uit de eerste lijn (bijvoorbeeld LINH) met die uit de tweede lijn (bijvoorbeeld LMR). Ook ziektespecifieke registraties bieden een schat aan gegevens. Koppeling van de kankerregistratie aan de GBA is nu nog niet mogelijk, maar lijkt een kwestie van tijd. Die koppeling zal het mogelijkheden maken om de grote sociaaleconomische verschillen in sterfte aan sommige kankersoorten te onderzoeken als functie van verschillen in kankerincidentie (en daarmee samenhangende leefstijlfactoren) en verschillen in kankeroverleving (en oncologische zorg).

### 3.3. *Mensen individueel volgen over de tijd*

Inherent aan de procesbenadering van het vorige punt is het feit dat individuele mensen longitudinaal over de tijd worden gevolgd. De mogelijkheid om elk persoon over de tijd te volgen, is een groot winstpunt van koppeling van registraties aan de GBA. Dit longitudinaal karakter biedt vooral nieuwe mogelijkheden voor onderzoek naar oorzakelijke relaties.

In veel onderzoek op het terrein van de volksgezondheid wil men statistische verbanden vaststellen voor zover die berusten op oorzakelijke relaties. Als men bijvoorbeeld het verband tussen overgewicht op sterfte wil onderzoeken, is men in de regel alleen geïnteresseerd in de verschillen die het effect van overgewicht op sterfte weerspiegelen. Omgekeerde effecten, bijvoorbeeld doordat het lijden aan ernstige ziekte resulteert in gewichtsverlies, zijn in zulke analyses alleen maar “verstoring”. Ook factoren als roken kunnen in zo’n analyse als “confounder” worden

beschouwd (Bonneux en Reuser, 2007). Om deze verstoring zo veel mogelijk te voorkómen, is een longitudinale opzet gewenst, waarbij overgewicht in de tijd wordt gemeten ruim vóór het optreden van ziekte en sterfte. Dit kan door bijvoorbeeld van respondenten in de POLS enquêtes het overgewicht ten tijde van het interview te meten en deze respondenten over de tijd te volgen wat betreft het optreden van gezondheidsuitkomsten zoals ziekenhuisopnamen of sterfte.

Soms zijn causale effecten in beide richtingen interessant. In bijvoorbeeld de gezondheidseconomie is men vaak geïnteresseerd in zowel de effecten van het genoten inkomen op gezondheid, als in de effecten van gezondheidsproblemen op het verdiend inkomen (van Doorslaer et al, 2008). In deze gevallen moeten causale effecten in twee richtingen worden onderzocht. Dan zijn complexere longitudinale designs gewenst. Men kan bijvoorbeeld gebruik maken van het feit dat gegevens uit fiscale bronnen beschikbaar zijn voor steeds meer jaren. Daardoor is het inkomen van een individu bekend in jaren vóór (vanaf circa 2000) en jaren na (tot 2010 of zelfs verder) het tijdstip waarop de gezondheid kan worden gemeten (bijvoorbeeld in de POLS 2005 enquête). Naarmate de te koppelen registers een langere tijdsperiode gaan bestrijken, zullen de mogelijkheden voor onderzoek naar complexe causale relaties groeien.

#### 3.4. *Bestudering van grote aantallen mensen*

Een belangrijk voordeel van integrale bestanden is hun grote statistische power. De omvang van gezondheidsproblemen, de grootte van ongelijkheden en de sterkte van statistische verbanden kan met grote nauwkeurigheid worden geschat. Illustratief zijn in dit verband de schattingen van sociaaleconomische verschillen in sterfte. Vroegere schattingen werden ontleend aan epidemiologische studies met 5,000 tot 20,000 respondenten (van Hertem et al, 2002). Hoewel deze aantallen in veel opzichten groot zijn, en een aanzienlijke investering in geld en menskracht vergden, leidden zij niet tot nauwkeurige schattingen van ongelijkheden in sterfte. In bijvoorbeeld het bekende GLOBE onderzoek was de verhouding in sterfte van laag opgeleiden tot hoog opgeleiden 1,64 met een 95% betrouwbaarheidsinterval (BI) van 1,13 tot 2,40 (Schrijvers et al, 1999). Kortom, de verschillen werden wel aangetoond (wat niet verrassend is) maar de precieze omvang van die verschillen bleef hoogst onzeker. Want volgens de BI was alleen waarschijnlijk dat de werkelijke omvang van het sterfteverschil ergens tussen de 13% en de 140% lag – een tienvoudig verschil!

De integrale schattingen die al in de jaren 1990 voor België konden worden gemaakt, gaven veel preciezere schattingen. Bijvoorbeeld de sterfte onder laag versus hoog opgeleide mannen van 60–69 jaar bedroeg 1,38 met 95% BI van 1,35 tot 1,41 (Huisman et al, 2004). Met deze gegevens werden niet alleen de verschillen aangetoond, maar kon ook de omvang van die verschillen nauwkeurig worden geschat. In die zin vertelden de integrale schattingen voor België veel meer nieuws dan de schattingen die in de jaren 1990 voor Nederland konden worden gemaakt. Gelukkig zijn nu voor Nederland wel integrale gegevens beschikbaar en zijn daarmee precieze schattingen gemaakt van de omvang van arm-rijk verschillen in levensverwachting.

Integrale gegevens bieden bovendien de mogelijkheid om onderscheid te maken tussen deelgroepen en/of deelgebieden die anders niet afzonderlijk onderzocht konden worden. In bijvoorbeeld een studie naar inkomensverschillen in ziekenhuisopnamen werd vastgesteld dat deze verschillen onder ouderen vrij klein waren, behalve als het ging om luchtwegaandoeningen. Ook onder 70 plussers was de kans om voor luchtwegaandoeningen opgenomen te worden veel groter onder mensen met een laag inkomen (Kunst et al, 2005). Dit resultaat suggereerde het bestaan van een, over de levensloop gezien, zeer persistente probleem. Luchtwegaandoeningen is onder ouder mannen in lagere sociale groepen mogelijk een groot probleem als gevolg van levenslange effecten van roken of van schadelijke blootstellingen vroeger op werk. Wat de verklaring ook is, het zou onmogelijk geweest zijn om deze verschillen onder ouderen te detecteren op basis van alleen steekproefonderzoek.

#### 4. *Mogelijkheden en beperkingen*

De vorige sectie liet zien welke mogelijkheden koppelingen bieden voor onderzoek op het terrein van de volksgezondheid. Maar uiteraard zijn die mogelijkheden niet onbegrensd. Waar liggen de grenzen?

Om te beginnen wil ik uiteen zetten waar op het terrein van de volksgezondheid registergegevens een belangrijke rol kunnen spelen. Op dit terrein kunnen drie typen onderzoek worden onderscheiden.

1. onderzoek dat tot doel heeft het vóórkomen van gezondheidsproblemen, ziekten en risicofactoren in kaart te brengen, met speciale aandacht voor ontwikkelingen over de tijd en voor verschillen tussen deelgroepen;
2. onderzoek dat tot doel heeft de beschreven ontwikkelingen en patronen te begrijpen, door kwantitatief onderzoek gericht op toetsing van oorzakelijke verbanden, en door kwalitatief onderzoek gericht op begrijpen van de onderliggende processen;
3. onderzoek dat tot doel heeft mogelijkheden voor interventies, zorg en beleid te identificeren, en de effecten van het gerealiseerde beleid te evalueren (vooraf of achteraf), mede door het opzetten van experimenten.

Het mag duidelijk zijn dat het koppelen van registers vooral mogelijkheden bieden voor het eerste type onderzoek, terwijl voor het tweede en het derde type onderzoek in mindere mate gebruik kan worden gemaakt van registergegevens.

De mogelijkheden en grenzen van onderzoek met registers wil ik in de rest van deze sectie illustreren aan de hand van een onderzoek dat wij nu in Amsterdam uitvoeren. Het betreft een onderzoek naar de gezondheid van bewoners van Vogelaarwijken (Droomers, 2009). In dit onderzoek beschrijven wij veranderingen in de gezondheidssituatie in Vogelaarwijken en in een groep van vergelijkbare achterstandswijken. Op basis hiervan beogen wij vast te stellen of in Vogelaarwijken sprake is van grotere verbetering in de gezondheidssituatie dan in de controlewijken.

ken. Indien inderdaad een sterkere verbetering wordt waargenomen, geeft dit enige "evidence" dat de verbeteringen van de algemene leefsituatie van de 40 Vogelaarwijken ook leiden tot een verbetering van de gezondheid van de bewoners. Dan betekent investeren in wijken ook investeren in gezondheid.

Maar hoe stellen we veranderingen in gezondheidssituatie in Vogelaar wijken vast? Neem het bijvoorbeeld van de sterfteregistratie. Deze biedt in principe grote mogelijkheden. De registratie is integraal, dekt alle buurten in Nederland en doet dit vanaf 1995. Daarmee kunnen veranderingen in de tijd worden vastgesteld in de sterftekansen van bewoners van Vogelaarwijken en controlewijken. Bovendien kan rekening worden gehouden met kenmerken van bewoners of huishoudens zoals het huishoudinkomen, land van herkomst en woningbezit. Tot slot is het mogelijk om rekening te houden met de verhuisgeschiedenis van individuele bewoners. Voor een beschrijvende analyse zijn deze gegevens van grote waarde, zeker als bovendien onderscheid wordt gemaakt naar doodsoorzaken. De belangrijkste beperking ligt in het relatief klein aantal sterfgevallen per buurt per jaar. Daardoor kunnen ontwikkelingen in sterfte waarschijnlijk niet met voldoende nauwkeurigheid geschat worden voor individuele wijken, maar alleen voor groepen van wijken.

Als door geringe statistische power sterfte niet als centrale uitkomstvariabele kan worden genomen, zijn bijvoorbeeld ziekenhuisopnamen een mogelijk alternatief. Het aantal ziekenhuisopnamen per jaar per buurt is vele malen groter dan het aantal sterfgevallen. Schattingen van de frequentie van ziekenhuisopnamen, op basis van de LMR, hebben daarmee een veel grotere nauwkeurigheid. Ook bij deze gegevens is bovendien onderscheid mogelijk naar ontslagdiagnose. Verder kan door koppeling aan de GBA ook in dit geval onderscheid worden gemaakt naar relevante kenmerken van bewoners en hun huishoudens, inclusief hun verhuisgeschiedenis. Met deze grote rijkdom aan gegevens is een proefschrift snel gevuld! Het voorbeeld van de LMR illustreert echter wel enkele problemen die in de praktijk vaak optreden bij koppeling van bestanden.

1. Vanwege het ontbreken van persoonlijke identificatienummers binnen de LMR is alleen een statistische koppeling met de GBA mogelijk. Een deel van de respondenten is daardoor niet uniek koppelbaar. Dit probleem treedt vooral op bij de oudere eerste generatie migranten (de Bruin et al, 2003).
2. Er zijn problemen als gevolg van het administratieve karakter van de LMR, zoals niet-deelname van klein aantal ziekenhuizen. Bovendien heeft een groot aantal ziekenhuizen recent besloten de gebruikelijke (op de internationale ICD classificatie gebaseerde) codering van ontslagdiagnosen te stoppen en alleen te coderen naar DBC (diagnose behandeling combinaties).
3. Praktische problemen zoals forse bedragen die de onderzoeker voor toegang tot de LMR bestanden aan de bestandhouder moet betalen. Door de complexiteit van de LMR bestanden zal de onderzoeker bovendien veel tijd moeten investeren in het bewerken en analyseklaar maken van bestanden.

Sommige van deze problemen zijn alleen praktisch van aard, maar andere proble-

men kunnen mogelijk leiden tot een vertekening van onderzoeksresultaten. Gezien al deze problemen zal een onderzoek met LMR gegevens niet snel worden begonnen, maar alleen indien daar veel tijd, geld en menskracht voor vrijgemaakt kan worden.

In het geval van ons onderzoek naar veranderingen in Vogelaarwijken is de keuze negatief uitgevallen. Analyses van de LMR registers of van de sterfteregistraties zouden vrijwel alle middelen opeisen, en ten koste gaan van andere deelonderzoeken. Daarnaast speelt een belangrijke inhoudelijke overweging. Het niveau van sterfte of de incidentie van de chronische aandoeningen reageert slechts langzaam op veranderingen in de leefsituatie van mensen. Vaak is er een vertraging van 10 jaar of meer, bijvoorbeeld als de effecten lopen via veranderingen in roken of overgewicht. Het is daarom niet waarschijnlijk dat de huidige veranderingen in de leefsituatie van Vogelaarwijken op korte termijn al zichtbaar zouden zijn in veranderingen in doodsoorzaken of ziekenhuisopnamen. Ook al zouden wij vinden dat ontwikkelingen in sterfte of ziekenhuisopnamen nu al gunstiger zijn in de Vogelaarwijken, dan nog is het niet aannemelijk dat die ontwikkelingen kunnen worden toegeschreven aan de nieuwe Wijkenaanpak.

Op basis van deze overwegingen zal een onderzoek als de onze, zoals in veel evaluatieonderzoek op het terrein van de volksgezondheid, zich bijvoorkeur richten op “intermediaire” factoren die sneller kunnen reageren op veranderingen en interventies. Daarbij kan worden gedacht aan risicofactoren die in het huidige preventiebeleid prioriteit hebben, zoals roken, slechte voeding, lichamelijke inactiviteit, en alcoholmisbruik. Ook “responsieve” indicatoren op het terrein van welzijn en geestelijke gezondheid kunnen relevant zijn, zoals het vóórkomen van angstgevoelens of van milde vormen van depressie.

Dit roept de vraag op: is het mogelijk om het vóórkomen van gedragsfactoren of welzijnsindicatoren in Vogelaarwijken te meten? En kan koppeling van registergegevens hierbij helpen? Hier stuiten we op een beperking van integrale registers: zij bevatten geen informatie over gedragsfactoren, of over ervaren gezondheid en welzijn. Dit is eenzelfde soort beperking als bij het opleidingsniveau van oudere generaties van Nederlanders: informatie hierover moet komen van persoonsenquêtes.

In geval van gedragsfactoren is een belangrijke bron van informatie de POLS enquêtes. Daarin worden jaarlijks een groot aantal respondenten gevraagd over verschillende gedragsfactoren. Ook van andere landelijke enquêtes is informatie over leestijl beschikbaar, zij het vaak over een beperkt aantal factoren. Bijvoorbeeld de WOON enquêtes van het CBS bevatten een vraag over lichamelijke activiteit. Het voordeel van WOON is daarbij dat zeer grote aantallen respondenten (meer dan circa 70.000) worden gevraagd, met in de laatste jaargang een oververtegenwoordiging van bewoners uit Vogelaarwijken (Blijie et al, 2010).

Andere mogelijkheden op dit terrein bieden bijvoorbeeld de vierjaarlijkse gezondheidsenquêtes van GGD'en, die steeds meer worden geharmoniseerd. Door



het RIVM wordt gewerkt aan een landelijke database op basis van GGD enquêtes. Een recente studie van het RIVM heeft op basis hiervan de gezondheidssituatie in 30 Vogelaarwijken in kaart gebracht (Ruijsbroek, 2010). De analyse toonde dat in deze wijken sommige gezondheidsproblemen meer voorkomen (bijvoorbeeld meer astma en diabetes) en dat ook roken en obesitas daar vaker voorkomt. Nader analyse wees uit dat die verschillen voor een deel kon worden verklaard door het gemiddeld genomen lagere opleidingsniveau van de bewoners van de Vogelaarwijken. Helaas bevatten de GGD enquêtes onvoldoende informatie over andere sociaal-economische factoren zoals inkomen en vermogen. In principe zou men door (statistische) koppeling met fiscale bronnen de GGD enquêtes op deze variabelen kunnen verrijken (Van Wingen et al, 2010). Verder onderzoek zou dan kunnen uitwijzen of de slechtere gezondheid van bewoners van Vogelaarwijken niet vooral het gevolg is van hun slechtere inkomens- en vermogenspositie. Concluderend, voor sommige soorten onderzoek zijn niet registers maar persoonsenquêtes de meest aangewezen bron. Maar ook hier kan koppeling met registergegevens een belangrijke rol vervullen, bijvoorbeeld door de enquêtegegevens te verrijken met sociaaleconomische gegevens van respondenten en hun huishoudens.

## 5. *Discussie*

### 5.1. *Over valkuilen*

In deze voordracht heb ik relatief weinig aandacht besteed aan “valkuilen” bij het gebruik van registergegevens. Alleen in vorige sectie heb ik bij wijze van illustratie opgesomd welke praktische problemen er spelen bij koppelingen met de LMR. Deze problemen zijn vergelijkbaar met de problemen bij de koppeling van de registers die in het SSB zijn opgenomen. Vergelijkbare problemen komt het CBS ook tegen wanneer koppelingen worden gemaakt met andere registers uit de zorgsector.

Ik wil op deze plaats mijn ervaring delen met een probleem die wij tegenkwamen bij gebruik van gegevens over inkomen uit fiscale bronnen (Kunst et al, 2005). Zoals mocht worden verwacht, vonden wij in een analyse van POLS gegevens een sterke relatie tussen de ervaren gezondheid van respondenten en hun inkomen volgens uit fiscale bronnen. Gezondheid konden wij ook onderzoeken in relatie tot het inkomen dat was gerapporteerd door de respondent zelf. Wij verwachtten de grootste arm-rijk verschillen in gezondheid te vinden bij het gebruik van geregistreerd inkomen, maar wij vonden de grootste verschillen in relatie tot het zelf gerapporteerd inkomen. Het feit dat inkomen uit fiscale bronnen minder voorspellend was, zou kunnen wijzen op problemen met de fiscale gegevens. Maar het probleem zou ook kunnen liggen bij de zelfrapportage van het inkomen: mensen die ertoe neigen hun eigen gezondheid slecht te beoordelen, zouden ook ‘somer’ kunnen zijn over hun eigen inkomen. Als dat zo is, zou dit probleem hebben geleid

tot een forse overschatting van arm-rijk verschillen in ervaren gezondheid – een overschatting waarvan wij jarenlang niet bewust waren! Een definitief antwoord kon niet worden gegeven. Wel illustreert dit dat koppeling van registergegevens ons onverwachte vragen kan stellen over de kwaliteit van gegevens – zowel aan de kant van de registers als aan de kant van de persoonsenquêtes.

### 5.2. *Over gebruik van registers voor onderzoek*

Vele mogelijkheden voor onderzoek openen zich als registers eenmaal zijn gekoppeld. Het SSB en het GSB zijn door een directeur van het CBS eens gekwalificeerd als een goudmijn. Dit is geen overstatement. Onze verkenning van de mogelijkheden van onderzoek naar doodsoorzaken of ziekenhuisopnamen in Vogelaarwijken illustreert dat die mogelijkheden talrijk zijn. Het is niet zinvol om aanbevelingen te doen over de keuze van de precieze onderwerpen. Het SSB en het GSB is wat dat betreft niet alleen een goudmijn, maar ook een snoepwinkel: er is voor elk wat wils. Wel is belangrijk dat elk onderzoek de methodische voordelen van registeronderzoek zo veel als mogelijk uitbuit. Die voordelen heb ik in sectie 3 uiteengezet. Uit dat overzicht kunnen de volgende aanbevelingen worden afgeleid:

waar dat informatief is, maak gebruik van de mogelijkheid om te differentiëren naar sociaaleconomische status, etnische herkomst of andere kenmerken van personen en hun huishoudens;

- maak gebruik van de mogelijkheden die gekoppelde registers bieden om uitkomstmaten te ontleden in samenstellende componenten (zoals sterfte in incidentie resp. case-fatality) of in fasen van het voorafgaand proces (zoals gebruik van tweede lijn in gebruik eerste lijn resp. doorverwijzing);
- maak bij de analyse van oorzakelijke verbanden gebruik van een longitudinaal design waarin de oorzaak wordt gemeten vóór het veronderstelde gevolg, of zet een longitudinal design op waarin “causality” en “reverse causality” van elkaar kunnen worden onderscheiden;
- benut de grote statistische power van integrale bestanden door vragen te beantwoorden naar de precieze omvang van veranderingen, ongelijkheden of effecten, of door onderscheid te maken tussen specifieke deelgebieden en/of deelpopulaties.

### 5.3. *Over uitbreiding van mogelijkheden voor registers*

De mogelijkheden voor onderzoek zullen groeien naarmate meer registerbestanden kunnen worden gekoppeld, de kwaliteit van bestaande registers verder wordt verbeterd, en personen over een langere tijd kunnen worden gevolgd. Het is daarom van onschatbare waarde dat het CBS blijft investeren in de kwaliteit van de registers en de koppelingen, en daarbij ruim aandacht besteedt aan valkuilen en problemen die de betrouwbaarheid van statistieken en onderzoeksresultaten kunnen schaden.

Niet minder belangrijk is dat nieuwe registers zullen worden toegevoegd aan het huidige systeem van koppelbare bestanden. Binnen gezondheidssector moet worden ge-

dacht aan landelijke registers voor specifieke sectoren (bijvoorbeeld thuiszorg en verpleging) en voor kosten van ziekten (gegevens van verzekeraars). Speciale aandacht verdienen registraties van de zorg in de eerste lijn, vanwege de vele soorten klachten en aandoeningen die hierin worden geregistreerd, in samenhang met informatie over ontvangen zorg in de eerste lijn en doorverwijzingen naar de tweede lijn ([www.nivel.nl/linh](http://www.nivel.nl/linh)). Ook koppeling aan ziekte-specifieke registraties, zoals de kankerregistratie, zou de mogelijkheid openen tot veel nieuw onderzoek. Een belangrijk voordeel van koppeling met ziektere REGISTRATIES is dat kenmerken van de patiënten en hun ziekteverloop kunnen worden onderzocht in relatie tot de ontvangen zorg ([www.ikcnet.nl](http://www.ikcnet.nl)).

Een belangrijke vraag is wat de plaats is van persoonsenquêtes waarbij respondenten slechts eenmaal, op één moment in de tijd, worden geïnterviewd. Deze enquêtes vormen de belangrijkste bron van informatie over: (a) risicofactoren, inclusief gedragsfactoren en psychologische kenmerken, (b) de ervaren gezondheid van mensen, inclusief algemeen welzijn en lichamelijke beperkingen, en (c) aspecten van zorggebruik die niet door bestaande registers worden gedekt. Ook kan worden gedacht aan baseline metingen van grootschalige epidemiologische studies, zoals het HELIUS onderzoek dat in 2010 van start zal gaan onder meer dan 50.000 Amsterdammers. Vanwege hun eenmalige en niet-integrale karakter zijn deze enquêtes en onderzoeken niet geschikt zijn voor “vulling” van het SSB of GSB, noch voor routinematige rapportages. Maar wel kan de omgekeerde weg worden gewandeld: door persoonsenquêtes of epidemiologisch onderzoek te koppelen aan registergegevens, zo niet prospectief dan wel retrospectief, kan aan dit onderzoek snel longitudinale diepte worden gegeven.

#### 5.4. Slot

Het overzicht in sectie 2 illustreerde dat onderzoek op het terrein van de volksgezondheid in Nederland is in de afgelopen 10 tot 15 jaar fundamenteel is veranderd. In dit opzicht schuift Nederland snel op in de richting van Scandinavische landen. Deze landen hebben zich lang van de rest van de wereld onderscheiden door de bijzondere mogelijkheden die koppeling van registers boden voor innovatief onderzoek. Het is binnen de epidemiologie gemeengoed dat, vanwege hun toegang tot die rijke databestanden, het handig is goede vrienden te zijn met collega's uit Scandinavië. Misschien valt ons die eer nu ook ten deel. Als het u opvalt dat buitenlandse collega's u de laatste jaren vriendelijker zijn gaan bejegenen, en u weet niet goed waarom, denk dan even aan het CBS.

## Referenties

Blijie, B., van Hulle, R., Poulus, C. en Hooimeijer, P. (2010). *Het wonen overwogen. resultaten van het WoonOnderzoek Nederland 2009*. Den Haag: Ministerie van VROM en het Centraal Bureau van de Statistiek.

Bonneux, L. en Reuser, M. (2007). Overgewicht en sterfterisico: geen samenhang tussen overgewicht op middelbare en oudere leeftijd en verhoogde sterfte. *Nederlands Tijdschrift voor Geneeskunde* 151, 2764–2769

Bos, V., Kunst, A.E. en Mackenbach, J.P. (2002). Sociaaleconomische sterfteverschillen in Nederland: een analyse op basis van buurtgegevens. *Tijdschrift voor Gezondheidswetenschappen* 80, 158–165.

Bos, V. (2005). *Ethnic Inequalities in Mortality in the Netherlands and the Role of Socioeconomic Status [proefschrift]*. Rotterdam: Erasmus Universiteit.

de Bruin, A., de Bruin, Gast, A., Kardaun J.W.P.F., van Sijl, M. en Verweij, G.C.G. (2003). *Koppeling van LMR- en GBA-gegevens: methode, resultaten en kwaliteitsonderzoek*. Voorburg/Heerlen: Centraal Bureau voor de Statistiek.

Droomers, M. (2009) Tackling health inequalities in the Netherlands. *Eurohealth* 15, 3, 16–18.

Garssen, J., Bos, V., Kunst, A.E. en van der Meulen, A. (2003). Sterftekansen en doodsoorzaken van niet-westerse allochtonen. *Bevolkingstrends* 3, 12–27.

Garssen, J. en van der Meulen, A. (2004). Perinatal mortality in the Netherlands – Backgrounds of a worsening international ranking. *Demographic Research* 11, 357–394.

Garssen, J., Harmsen, C. en de Beer, J. (2005) The effect of the summer 2003 heat wave on mortality in the Netherlands. *Eurosurveillance* issue 7, article 7.

Huisman, M., Kunst, A.E., Andersen, O., et al. (2004). Socioeconomic inequalities in mortality among elderly people in 11 European populations. *J Epidemiol Community Health* 58, 468–75.

Janssen, F., Mackenbach, J.P. en Kunst, A.E. (2004). Trends in old-age mortality in seven European countries, 1950–1999. *J Clin Epidemiol* 57, 203–216.

Kardaun, J.W.P.J. en Glerum, J.J. (1995). *Sterfte, doodsoorzaak en inkomen*. Voorburg/Heerlen: Centraal Bureau voor de Statistiek.

Klijs, B., Mackenbach, J.P., Nusselder, W. en Kunst, A.E. (2010). *Risk factors for spending life years with disability: Sullivan life table calculations including time to death [manuscript]*. Rotterdam: Erasmus MC.

Knoops, K. en van den Brakel, M. (2010). Rijke mensen leven lang en gezond. Inkomensgerelateerde verschillen in de gezonde levensverwachting. *Tijdschrift voor Gezondheidswetenschappen* 88, 17–24.

Kunst, A.E., Mackenbach, J.P., Lautenbach, H., Oei, F.B. en Bijlsma, F. (2003). Gains in life expectancy by eliminating major causes of death: revised estimates taking into account competing causes of death. In: Wunch, G. et al (ed) *The life table: modelling survival and death*. Dordrecht: Kluwer Academic Publishes, p. 191–206.

Kunst, A.E., Dalstra, J.A., Bos, V., Mackenbach, J.P., Otten, F.W.J. en Geurts, J.J. (2005). *Ontwikkeling en toepassing van indicatoren van sociaal-economische status binnen het Gezondheidsstatistisch Bestand*. Voorburg: Centraal Bureau voor de Statistiek.

Kunst, A.E. (2007). Sociaal-economische verschillen in sterfte en gezondheid in Nederland. Een overzicht van vorderingen in recent beschrijvend onderzoek. *Bevolkingstrends* 55: 34–44.

Kunst, A.E. en Droomers, M. (2009). *Evaluatie van schattingen van sociaal-economische verschillen in de gezonde levensverwachting in Nederland*. Amsterdam: AMC.

Lau-IJzerman, A., Habbema, J.D.F., Maas, P.J. van der, et al. (1981). *Vergelijkend buurtonderzoek Amsterdam*. Amsterdam: GGenGD.

Mackenbach, J.P. en Stronks, K. (2004). The development of a strategy for tackling health inequalities in the Netherlands. *International J Equity and Health*. 3, 11.

Mackenbach, J.P. en van der Maas, P. (2008). *Volksgezondheid en gezondheidszorg*. Amsterdam: Reed business.

Mackenbach, J.P., Stirbu, I., Roskam, A.J., Schaap, M.M., Menvielle, G., Leinsalu, M. en Kunst, A.E. (2008). Socioeconomic inequalities in health in 22 European countries. *New England J Medicine* 358, 2468–2481.

Ruijsbroek, A., Droomers, M., Kunst, A.E. en van den Brink, C. (2010). Gezondheid en leefstijl in de krachtwijken: een verkenning. *Tijdschrift voor Gezondheidswetenschappen*, in bewerking.

Schrijvers, C.T., Stronks, K., van de Mheen, H.D. en Mackenbach, J.P. (1999) Explaining educational differences in mortality: the role of behavioral and material factors. *American J Public Health*. 89, 535–40

Stam, S. Garssen, M.J., Kardal, M. en Lodder, B.J.H. (2008). Hoogopgeleiden leven lang en gezond. In: van Hilten, O. en Mares, A.M.H.M. (red) *Gezondheid en zorg in cijfers 2008*. Den Haag/Heerlen: Centraal Bureau voor de Statistiek.

Stirbu, I., Kunst, A.E., Vlems, F.A., Visser, O., Bos, V., Deville, W. en Coebergh J.W.W. (2006). Cancer mortality rates among first and second generation migrants in the Netherlands: Convergence toward the rates of the native Dutch population. *International J Cancer* 119, 2665–2672.

Stirbu, I. en Mackenbach, J.P. (2009). *Ongelijkheid in gezondheid, is gezondheidszorg van belang? Sociaaleconomische en etnische verschillen in gezondheidszorguitkomsten op het terrein van hart- en vaatziekten in Nederland*. Utrecht: Inspectie voor de Gezondheidszorg.

van Beeck, E.F., Mackenbach, J.P., Looman, C.W.N. en Kunst, A.E. (1991). Determinants of traffic accident mortality in The Netherlands: a geographical analysis. *International J Epidemiology* 20, 698–706.

van Doorslaer, E., van Kippersluis, H., O'Donnell, O. en Van Ourti, T. (2008). *Socio-economic differences in health over the life cycle: evidence and explanations*. Netspar Panel paper 12. Tilburg: Printing Office Tilburg University

van der Lucht, F. en Polder, J.J. (2010). *Van Gezond naar beter. Volksgezondheid Toekomst Verkenning 2010*. Houten: Bohn Stafleu Van Lochem.

van Herten, L.M., Oudshoorn, K., Perenboom, R.J.M., Mulder, Y.M., Hoeymans, N., Deeg, D.J.H., Kunst A.E. en van Lenthe F. (2002). *Gezonde levensverwachting naar sociaal-economische status*. Leiden: TNO.

van Poppel, F. en Joung, I. (2001) Long-term trends in marital status mortality differences in The Netherlands 1850–1970. *J Biosocial Science* 33, 279–303.

van Wingen, M., Berger-Van Sijl, M., Kunst, A.E. en Otten, F. (2010). Inkomen en vermogen als indicatoren van gezondheidsverschillen. *Bevolkingstrends* 58, 43–49.

# *Continuïteit van zorg: Indicatoren voor lange termijn zorg en zekerheid in de GGz*

André Wierdsma (Erasmus MC)

## *1. Landelijke en regionale GGz-informatie*

In de geestelijke gezondheidszorg (GGz) zijn grote logistieke problemen opgelost: in de kliniek kan voor iedere patiënt precies op de juiste plaats en tijd een maaltijd volgens een bijzonder dieet worden geserveerd. Er wordt echter onvoldoende geregistreerd of het met patiënten ook echt beter gaat. Deze merkwaardige tegenstelling constateerde de huidige directeur van het Sociaal en Cultureel Planbureau, Paul Schnabel, in zijn "De weerbarstige geestesziekte". Het boek verscheen in 1995, maar het commentaar geldt in grote lijnen nog steeds. Op landelijk niveau en in de regionale samenwerking van voorzieningen zijn GGz-monitors ontwikkeld vaak met als doel de veranderingen in de zorg te evalueren. Toch laat de informatievoorziening over de kwaliteit van de GGz nog veel te wensen over. Zowel nationaal als regionaal was het informatiebeleid de afgelopen decennia vooral 'trial and error'.

De oudste landelijke registratie is het Patiëntenregister Intramurale GGz (PIGG) van de psychiatrische ziekenhuizen. Met de extramuralisering van de zorg kwam daar het RIS/NIS bij: de registraties van de Regionale Instellingen voor Ambulante GGz (RIAGG informatie systeem), gebundeld in een database onder beheer van de koepelorganisatie (NVAGG informatie systeem). De verslavingszorg kende een vergelijkbare gelaagde informatievoorziening: CADIS/LADIS (Consultatiebureau respectievelijk Landelijk Alcohol en Drugs informatie systeem). In praktijk was de waarde van deze nationale informatie beperkt omdat de overlap van de zorg niet in beeld werd gebracht, een substantieel deel van de instellingen niet participeerde en de kwaliteit van de aangeleverde data soms twijfelachtig was (Rijter et al., 2002). De start van ZORGis in 2000 voegde de landelijke informatiesystemen samen. Daartoe was een kerngegevensset opgesteld met expliciete eisen ten aanzien van de classificatie en codering van de zorginformatie en het format van de gegevensaanlevering. In de eerste jaren kende ZORGis onder de deelnemende instellingen echter een respons van niet meer dan rond de 50%. Bovendien werden bij sommige items, bijvoorbeeld aanmeldingsklacht en diagnose, veel ontbrekende waarden ingevuld. Na 2007 is deze database niet voortgezet en is alle aandacht gericht op het Diagnose-Behandel Combinatie Informatie Systeem (DIS). Vanaf 1 januari 2006 zijn GGz-instellingen en vrijgevestigde praktijken wettelijke verplicht om de productie in DBC-terminen te registreren en bij het DIS aan te leveren. Maar niet alle aanbieders waren bij de start al

gereed om met aanlevering te beginnen. Bovendien doet in deze systematiek een belangrijk deel van de openbare GGz niet mee, namelijk de zorg voor chronische patiënten in de verblijfspsychiatrie en in de maatschappelijke opvang.

Het Ministerie van VWS gaf niet alleen steun aan deze landelijke ontwikkelingen maar ook aan regionale GGz-registers. Eind jaren tachtig werd het plan opgevat om op termijn het oude PIGG te vervangen door een samenwerkingsverband van regionale registers (ten Horn, 1989). Naar het voorbeeld van de eerste psychiatrische casusregisters in Engeland zijn registers opgezet voor Noord-Nederland, Zuid-Limburg, en Rotterdam-Rijnmond; een vierde register voor Utrecht en omstreken is sinds enkele jaren actief (Giel & Sturmans, 1996). Door de informatiseringsgolf ontstond echter een wirwar aan regionale systemen waarin het Ministerie begin jaren negentig enige lijn probeerde te brengen (Vermande & Bijl, 1995). De aan universitaire afdelingen verbonden casusregisters zouden de peilers zijn in de verdere ontwikkeling van de mogelijkheden en het gebruik van de zorggegevens per regio. Een Landelijk Platform Informatievoorziening werd ingesteld om te komen tot een dekkende verzameling van RIPAG's: Regionale Informatiesystemen Patiëntgegevens GGz. Enkele jaren geleden is deze ontwikkeling echter stopgezet. Debet aan de RIPAG-stop waren de vele fusiebewegingen waardoor de noodzakelijke veranderingen in de registraties van GGz-instellingen prioriteit kregen.

De beperkingen van landelijke databestanden maken duidelijk dat de bestaande psychiatrische casusregisters samen een unieke verzameling van gegevens bieden die de ontwikkeling van de GGz in beeld kunnen brengen. Maar de opbouw van deze gegevens laat ook de complexiteit zien van registeronderzoek naar verschillen tussen regio's en ontwikkelingen in de tijd.

## 2. *Mogelijkheden en valkuilen van registeronderzoek*

In een psychiatrisch casusregister worden de contacten met GGz-instellingen van patiënten uit een omschreven geografisch gebied op één centraal punt verzameld en gekoppeld. De gegevensset omvat patiëntinformatie, zorgepisodeinformatie, en contactgegevens. Patiënten kunnen in de tijd en over verschillende zorgvoorzieningen worden gevolgd (ten Horn, 1989). Registeronderzoek gaat in op trends in het zorggebruik, risicofactoren, bijzondere patiëntgroepen, en continuïteit van zorg (Wierdsma et al., 2007).

### 2.1 *Trends in zorggebruik: prevalentie en incidentie studies*

De registers geven samen een representatief beeld van de effecten van zorgbeleid en de regionale verschillen. Het extramuraliseringsbeleid in de jaren '90 heeft er toe geleid dat opnamen in de psychiatrische kliniek steeds vaker worden vervangen door alternatieve zorgvormen, zoals dagbehandeling en beschermd wonen. De praktijk verschilt echter per regio (Pijl et al., 2005; Wierdsma, Pijl et al., 2006). Er



zijn ook regionale verschillen in de toepassing van dwang wanneer een patiënt een gevaar vormt voor zichzelf of de omgeving. De omvang en de diversiteit van de problematiek zijn het grootst in stedelijke gebieden (Wierdsma et al., 2010).

Eén van de valkuilen in dergelijke prevalentie- en incidentiestudies is het toespitsen van registergegevens op relatief zeldzame stoornissen zonder rekening te houden met regionale verschillen in de organisatie van de zorg. Blansjaar (1992) schatte het prevalentiecijfer van Korsakoff-patiënten in Den Haag op bijna 4,8 per 10 000 inwoners. Voor de provincie Noord-Holland kwam Schnabel (1992) ongeveer een derde lager uit: rond 3 per 10 000. Blansjaar's cijfer is vermoedelijk een overschatting omdat gespecialiseerde Haagse voorzieningen een regionale functie hadden. Niet het aantal inwoners van de stad, maar het aantal van een wijder gebied had mogelijk in de epidemiologische noemer moeten staan (Wierdsma & Sytema, 1996).

## 2.2 *Risicofactoren*

Het onderzoek van risicofactoren op basis van administratieve gegevens is beperkt tot indicatoren van determinanten van psychiatrische stoornissen die in de dataset zijn meegenomen, bijvoorbeeld leeftijd, geslacht en geboorteland. Etniciteit als risicofactor is een belangrijk thema in registeronderzoek: mannelijke migranten uit de Nederlandse Antillen of uit Marokko hebben een verhoogd risico op schizofrenie en ook op middelenafhankelijkheid (Selten et al., 2007). Reviews van etnische verschillen in gedwongen opnamen laten een heterogeen beeld zien, waarbij de relatieve risico's hoger uitkomen in studies die niet corrigeren voor relevante factoren (Singh et al., 2007). Mulder et al. (2006) schatten het risico op gedwongen opnamen voor Surinamers en Antillianen ruim drie keer zo hoog als voor Nederlanders (gecorrigeerd voor leeftijd en geslacht). Daarbij is echter uitgegaan van de bevolking in de regio Rijnmond, terwijl etnische minderheden vooral wonen in de Rotterdamse achterstandswijken. Daardoor is het effect van etniciteit vermengd met verstedelijking en sociaal-economische gezondheidsverschillen.

## 2.3 *Bijzondere patiëntgroepen*

Bijzondere groepen worden onderscheiden op basis van patiëntkenmerken, de inschrijvingsduur en de aard en intensiteit van de hulpverlening. Dergelijk onderzoek geeft inzicht in de diversiteit van de GGz-doelgroep en benodigde zorgprogramma's. Daaruit blijkt bijvoorbeeld dat langdurig zorgafhankelijke patiënten 15% vormen van de populatie in zorg, maar deze betrekkelijk kleine patiëntengroep krijgt ongeveer 66% van de intensieve zorg en 33% van de ambulante hulpverlening (Dieperink et al., 2006). Een valkuil bij dit type onderzoek is onbekendheid met de regionale context van zorggegevens. Scholten en Tjadens (1996) rapporteerden dat ongeveer 25% van de gedwongen psychiatrische spoedopnamen niet door de rechter wordt voortgezet. Rotterdam voerde de lijst aan met maar liefst 42%. Vermoedelijk zijn patiënten meegeteld die buiten de regio werden geplaatst. De juridische procedure vereist

dan dat i.c. de Rotterdamse rechtbank de dwangopname niet bevestigt en de beoordeling overlaat aan de rechtbank in het gebied waar de patiënt is opgenomen. Een beddentekort in Rotterdam zorgde in die tijd voor een relatief groot aantal zogenoemde gastplaatsingen.

#### 2.4 *Continuïteit van zorg*

In de psychiatrische epidemiologie was Sineke ten Horn waarschijnlijk de eerste die het onderwerp zorgcontinuïteit heeft opgepakt in haar proefschrift 'Nazorg geeft kopzorg' (ten Horn, 1982). Het thema is nog altijd actueel: voor bijna één op de vijf gedwongen opgenomen patiënten is er geen passende nazorg (Stobbe et al., 2009; Wierdsma, van Baars et al., 2006). GGz-voorzieningen werken aan transparantie van de zorg onder andere met zogenoemde prestatie-indicatoren. De enige indicator in de werkinstructie voor 2009–2010 die betrekking heeft op de nazorg is: "Het percentage cliënten dat na ontslag uit de kliniek binnen twee weken een ambulante of deeltijdcontact heeft bij de zorgaanbieder waar de klinische opname heeft plaatsgevonden." Maar dit is een erg magere operationalisatie van continuïteit van zorg. Een valkuil in het gebruik van registergegevens is het ontbreken van een conceptueel kader waardoor het onderzoek vooral 'data-drijven' blijft.

In de geestelijke gezondheidszorg (GGz) draagt continuïteit van zorg bij aan betere diagnostiek, meer therapietrouw, minder heropnamen en kortere opname duur (Haggerty et al., 2003; Reid et al., 2002). Daarom staat zorgcontinuïteit in veel landen hoog op de maatschappelijke en politieke agenda. Ook in Nederland is door de Gezondheidsraad het belang onderkend van continuïteit in de openbare GGz. In dit zorgveld gaat het om ernstig zieke en vaak langdurig zorgafhankelijke patiënten die niet zelf om hulp vragen. Het adviesrapport over de acute psychiatrische zorg stelt dat zorgcontinuïteit noodzakelijk is om te voorkomen dat patiënten tussen de wal en het schip raken (Gezondheidsraad, 2004). Dit komt echter niet tot uitdrukking in de prestatie-indicatoren. De geringe aandacht voor de nazorg komt onder andere door het ontbreken van standaard maten van zorgcontinuïteit aan de hand van administratieve data (Wolinsky et al., 2007). Daarom worden hierna op basis van de literatuur en analyses van registratiegegevens, de mogelijkheden en moeilijkheden van het monitoren van zorgcontinuïteit geïnventariseerd.

### 3. *Indicatoren van zorgcontinuïteit*

Continuïteit van zorg betekent dat opeenvolgende zorgepisodes naadloos met elkaar worden verbonden in een zorgpatroon dat aansluit op de behoeften van de patiënt (Bachrach, 1981). In de literatuur zijn vier typen zorgcontinuïteit te onderscheiden: (a) informatie-overdracht tussen zorgaanbieders, (b) coördinatie van het

zorgaanbod bijvoorbeeld in 'disease management' programma's, (c) de ontwikkeling van persoonlijke relaties, en (d) flexibiliteit van contacten of contactfrequentie op geleide van de behoefte van de patiënt (Haggerty et al., 2003; Wierdsma et al., 2009). Zorgcontinuïteit is een ingewikkeld concept omdat verschillende disciplines eigen accenten leggen, bijvoorbeeld in de eerstelijnszorg weegt het opbouwen van een vertrouwensband zwaar. Ook kunnen verschillende niveaus van het zorgsysteem worden onderscheiden. Continuïteit van zorg heeft betrekking op het uitwisselen van patiënt informatie op regionaal niveau, het aanbieden van een continuüm van ambulante en klinische zorg op instellingsniveau, een interdisciplinaire aanpak in het behandelteam en flexibiliteit van contacten op patiënt niveau. In het jargon van multilevelanalyse heeft continuïteit van zorg een vier-level-structuur: observaties zijn geclusterd binnen patiënten en patiënten zijn geclusterd binnen zorgaanbieders en gezondheidsregio's.

Beter zicht op de wisselwerking tussen de verschillende niveaus levert informatie over de mechanismen om continuïteit van zorg te verbeteren. Vaak wordt daarbij het patiëntperspectief centraal gesteld (Freeman et al., 2002). De patiënt overziet echter niet het volledige zorgproces waardoor subjectieve maten een weinig gedifferentieerd beeld kunnen opleveren (Gulliford et al., 2006). De ervaren continuïteit kan een belangrijke kwaliteitsindicator zijn in tevredenheidsonderzoek onder patiënten, maar administratieve maten zijn belangrijk in de evaluatie van continuïteit van zorg (Carr et al., 2003). Read et al. (2002) waarschuwen echter voor het gebruik van administratieve data in indicatoren van zorgcontinuïteit omdat met registratiegegevens niet duidelijk wordt welke typen van continuïteit worden gemeten. Om dat te onderzoeken, hebben wij indicatoren geïnventariseerd en is op basis van registergegevens nagegaan welke typen van zorgcontinuïteit te onderscheiden zijn.

### 3.1 *Inventarisatie van continuïteitsmaten*

Het literatuuronderzoek richtte zich op Engelstalige publicaties op het terrein van de medische behandeling en gezondheidszorg, verpleegkunde en verwante bronnen (Medline, PubMed, SCOPUS). De volgende zoektermen werden gebruikt: 'continuity', 'patterns of care', 'aftercare', 'use of mental health care', in combinatie met 'measures', 'measurement', 'index', 'indices' of 'indicators'. Referentielijsten zijn doorgenomen op aanvullende artikelen of rapporten en de eerste verwijzingen naar het gebruik van de continuïteitsmaten in gezondheidszorgonderzoek.

Voor de samenvatting van het verloop van de zorg over een bepaalde observatieperiode zijn 15 indicatoren gevonden, op alfabetische volgorde opgenomen in tabel 1 met een samenvatting van de betekenis en operationalisering. Hierna wordt voor de leesbaarheid gebruik gemaakt van de afkortingen. Per indicator is een korte beschrijving gegeven en de wijze van berekening met de range van de uitkomsten. Tevens zijn de waarden van het rekenkundig gemiddelde en de mediaan vermeld op basis van de registratiegegevens die in het onderzoek zijn gebruikt (zie hierna).

**Tabel 1**  
**Continuïteit-van-zorg maten met administratieve data, in alfabetische volgorde: omschrijving en scores berekend voor een**  
**incidentie-cohort over een 30 maanden follow-up (patiënten die minimaal één keer werden opgenomen in de observatieperiode;**  
**N=1554)**

Index	Eerste referentie	Formule / range	Cohort gemiddelde (Mn) en mediaan (Mdn)
CFC	Classification of Fragmentation of Care: high, moderate or low classification based on the number of distinct hospitals (i.c. type of care) a patient visited relative to patients' total number of hospitalisations (i.c. number of intakes). The classification is replicated for ascending numbers of admissions or intakes.		
	Schrag et al. (2006)	Three admissions in 1 hospital: low fragmentation, in 3 hospitals: high fragmentation; six admissions in 1 or 2 hospitals: low, in 5 or 6 hospitals: high fragmentation of care; etcetera. Range 1 to 3	Mn= 1.41 (.555)  Mdn= 1.0 Range 1 - 3
Cla	Alpha Index of COC: combines concentration of providers and sequential continuity in a series of face-to-face contacts. The concept of Kullback-Leibler Information is used to represent the degree of concentration relative to no concentration at all, and SECON is used as a measure of sequential continuity. The parameter alpha specifies the pre-determined weight that is given to the distinct dimensions (i.c. 0.5). The number of available providers is defined here as the maximum number of providers that any one patient has visited (i.c. number of healthcare services).		
	Lou (2000)	$Cla = a KL^* + (1-a)SECON$ , where $KL^* = KL(P(n), Q_m) / \log m$ , n=number of visits, m= number of available providers Range 0 to 1	Mn= .89 (.141) Mdn= .968 Range .37 - 1.0
COCI	Continuity-of-Care Index: the patient's total number of contacts for a specific time period relative to the number of providers or groups of referred providers (considering all contacts to one's usual provider or any referred providers as contacts with a single provider - i.c. inpatient and outpatient services in the same sub-regions were grouped).		
	Bice & Boxerman (1977)	$SUM((\#contacts\ with\ provider)^2 - \#contacts) / (\#contacts(\#contacts - 1))$ Range 0 to 1	Mn= 0.89 (.196) Mdn= 1.0 Range .04 - 1
CON	Standardized Index of Concentration: where the number of providers (or categories of referral) differs between areas or is likely to shift over time, a standardized index is needed. This index of concentration (i.c. per healthcare services) equals 0 when the proportion of contacts is the same for all categories and 1 when 100% is concentrated in one category.		
	Shortell (1976)	$SQR(HHi \div (1 - 1/\#providers) / (1 - 1/\#providers))$ Range 0 to 1	Mn= 0.92 (.122) Mdn= 0.988 Range 0 - 1
DOR	Duration of Relationship (Patient/Provider Affiliation): length of time from initial to final contact between patient and provider, e.g. more or less than five years. Provider is the usual healthcare professional or treatment team (i.c. healthcare services).		
	McWhinney et al. (1988)	0 to $\infty$	Mn= 531.3 (335.9) Mdn= 614 Range 1 to 900
FRAC	Fraction of care index (contact proportion) is a measure of the share of providers, or sites, relative to the total number of contacts - in economic studies used as indicator of the amount of competition among firms (i.c. healthcare services).		
	Roos et al. (1980)	$1 - HHi \div$ Range 0 to 1	Mn= 0.16 (.226) Mdn= 0.034 Range 0 - 1
GINI	The Gini index measures the extent to which the distribution of contacts among types of care within an observation period deviates from a perfectly equal distribution (the area between the Lorenz curve and the hypothetical line of absolute equality). A Gini coefficient of zero represents an equal and 1 an unequal distribution (i.c. per healthcare service).		
	Shortell (1976)	$1 - SUM(Yi+1 + Yi)(Xi+1 - Xi)$ Range minus $1-1/n$ to plus $1-1/n$	Mn= 0.29 (.270) Mdn= 0.348 Range 0 - .75

Tabel 1 (slot)

Continuïteit-van-zorg maten met administratieve data, in alfabetische volgorde: omschrijving en scores berekend voor een incidentie-cohort over een 30 maanden follow-up (patiënten die minimaal één keer werden opgenomen in de observatieperiode; N=1554)

Index	Eerste referentie	Formule / range	Cohort gemiddelde (Mn) en mediaan (Mdn)
HSM	Hierarchical Stage Model combines the intensity of care and level of care into one classification (here primary care data were not available). If a patient receives several types of care, the sum of corresponding scores represents a unique combination of intensity and type of care. Ten Horn & Haffner (1989)	Every patient receives a score once every 6 months, 0 = no treatment, (1-2 = primary care), 3-6 = outpatient, 9-18 = inpatient, including day care Range 0 to 26 (per 6 months)	Mn= 23.8 (14.4)  Mdn= 24 Range 3 – 75 (cumulative over five periods)
ICI	Intensity of Care index (Frequency of care): the proportion of the gross care duration (number of days between the first and the last contact in the observation period) in which someone was actually receiving care; no-care period is defined as longer than 90 days without care. Sytema et al. (1985)	(Gross duration – No-care periods) / Gross duration	Mn= 0.69 (.309) Mdn= .766 Range 0 – 1
KPC	Known Provider index (K-index; Ejlertsson's-index) is the number of contacts with a professional, treatment team or provider (i.e. type of care per healthcare service) seen previously relative to the total number of contacts. The smaller this difference is, the better continuity of care. Ejlertsson & Berg (1984)	(#contact s – # providers) / (#contacts – 1) Range 0 to 1	Mn= 0.96 (.101) Mdn= 0.986 Range 0 – 1
MMCI	Modified-Modified Continuity Index (Consistency measure) is a measure of concentration of care with providers, or over facilities, derived from MCI: $1 - (\#providers / (\#contacts+0.1))$ to produce a range in scores from 0 to 1 (i.e. indicating concentration of contacts per healthcare service). Magill & Senf (1987)	$(1 - (\#providers / (\#contacts+0.1))) / (1 - (1/(\#contacts+0.1)))$ Range 0 to 1	Mn= 0.98 (.066) Mdn= 0.997 Range .09 – 1
NOI	Number of providers or locations (i.e. types of care per healthcare service) with whom the patient had contact during the observation period. Provider varies from integrated mental healthcare organisations to specialized psychiatric services. Hall et al. (1994)	Providers: 0 to ∞	Mn= 3.26 (1.83) Mdn= 3.0 Range 1 to 17
SCi	Service Consumption index is a cumulative index for service consumption mixing the time element with level of care, based on weight points for out-patients consultation (1), days in day-patient service (2) and in in-patient service (3). Lavik (1983)	1 to maximum (3*observation period)	Mn= 366.2 (416.6) Mdn= 219.5 Range 1 to 2505
SeCon	Sequential Continuity is a measure of the sequential nature of provider continuity that equals the fraction of sequential pairs of contacts at which the same provider is seen (i.e. types of care per healthcare service); expected value is 1 /number of available providers if care seeking patterns are random. Steinwachs (1979)	(#sequential contacts / #contacts-1) Range 0 to 1	Mn= 0.88 (.146) Mdn= .919 Range 0 to 1
UPC	Usual Provider Continuity (Most Frequent Provider Index): proportion of visits to the usual source of care. Usual provider can be operationalized as the provider with the largest number of contacts or the patient's preferred provider (i.e. per healthcare service). Breslau & Reeb (1975)	(#contacts with provider / #contacts) Range 0 to 1	Mn= .90 (.146) Mdn= .988 Range .33 – 1

‡ Herfindahl-Hirschman Index:  $SOM((\#contacts\ with\ I / \# all\ contacts)^2)$ , waarbij I staat voor behandelaar, zorginstelling of verwijzer, range 0 tot 1

In dit overzicht zijn enkele andere benaderingen buiten beschouwing gelaten, bijvoorbeeld indicatoren die uitgaan van veranderingen in de tijd, zoals de tijdsduur tussen ontslag en vervolcontact in survival analyses (Sytema et al., 2002) of de waarschijnlijkheid van verandering van zorgtype in Markov-keten analyses (Moreno et al., 2007). Enkele indicatoren zijn weggelaten die andere continuïteitsmaten combineren of aanpassen voor een bijzondere setting. De 'Fundamental Continuity of Care Index' combineert bijvoorbeeld de UPC index: gebruikelijke zorgaanbieder, en SCi-score: de cumulatieve duur van de begeleiding (Citro et al., 1997). Soms worden geobserveerde scores afgezet tegen de waarden wanneer een random verdeling kan worden verwacht (Steinwachs, 1979). Een andere benadering is verklaringsmodellen van continuïteit van zorg te vergelijken, bijvoorbeeld wisselingen van zorgaanbieder, rekening houdend met relevante kenmerken van patiënten en zorgaanbieders in plaats van random zorgpatronen (Hussein & Carrière, 2002). Andere indicatoren die in het overzicht niet zijn meegenomen, vereisen aanvullende informatie om waargenomen en verwachte uitkomsten te vergelijken. De 'Temporal Continuity Index' gaat bijvoorbeeld uit van de tijdsintervallen tussen zorgcontacten in verhouding tot de optimale follow-up, wat alleen kan worden toegepast wanneer er een algemeen geaccepteerde standaard is (Spoonner, 1994). Tot slot zijn continuïteitsmaten buiten beschouwing gelaten die in sterke mate afhangen van coderingen of de administratieve discipline, zoals bijvoorbeeld contacten met zorgcoördinatoren of verwijzingen vanuit bepaalde samenwerkingsverbanden (Brown et al., 1994; Lehman et al., 1994).

### 3.2 *Vormen van zorgcontinuïteit*

Met gegevens van het psychiatrisch casusregister voor Rotterdam-Rijnmond is gekeken naar de samenhang van indicatoren van zorgcontinuïteit. De continuïteitsmaten zijn berekend voor een incidentiecohort van patiënten die 30 maanden in zorg bleven en contact hadden met zowel ambulante als klinische of deeltijd voorzieningen. Als eerste stap werd een cohort geselecteerd van patiënten die in de jaren 1997 tot 2001 een eerste GGz-contact hadden (N=52095). Vervolgens werden patiënten van het onderzoek uitgesloten die 2 ½ jaar na het eerste contact niet meer in zorg waren (N=12983), zodat voor alle patiënten in het onderzoekscohort een gelijke observatieperiode van 30 maanden kon worden aangehouden. Omdat we contacten over verschillende typen zorg wilden volgen, werden ook de patiënten verwijderd die in de observatieperiode niet zijn opgenomen (N=37096) of langer dan 12 maanden in een kliniek verbleven (N=99). Tot slot is de onderzoeksgroep beperkt tot patiënten in de leeftijd van 15 tot 60 jaar, omdat van de gespecialiseerde zorgcircuits voor kinderen en ouderen niet alle zorginstellingen aan het casusregister deelnemen. Van alle patiënten (N=1554) zijn de klinische en ambulante contacten geselecteerd over een follow-up periode van 30 maanden na het eerste contact met een van de GGz-instellingen.

Indicatoren van zorgcontinuïteit werden berekend en de dimensionaliteit van continuïteit in de geestelijke gezondheidszorg is onderzocht met een explorerende factor-

**Tabel 2**  
**Principale factor analyse van continuïteit van zorgindicatoren; communaliteiten en structuurmatrix na oblimin rotatie**  
**(drempelwaarde .40)**

Indicatoren *	Factor 1 Informatie continuïteit	Factor 2 Contact continuïteit	Factor 3 Relatie continuïteit	Factor 4 Management continuïteit	Communaliteit
<i>Concentratie</i>					
CIA	,967			-,608	,951
COCI	,643			-,571	,489
CONI	,990			-,582	,989
MMCI	,675	,452		-,669	,711
UPC	,989			-,572	,987
<i>Fragmentatie</i>					
CFC	-,480			,697	,501
FRAC	-,993			,580	,995
GINI	-,664				,445
<i>Sequentieel</i>					
SECON			,554		,316
KINDEX		,594	,659		,788
<i>Frequentie</i>					
DOR					,221
HSM		,857			,778
ICI		,402			,176
NOI			-,518	,436	,463
SCI		,682			,497
Eigenvalue	6,400	2,294	1,348	0,89	
% Totale variantie	42,7%	58,0%	66,9%	72,9%	

\* Zie Tabel 1 voor de afkortingen.

analyse. Alle bewerkingen en analyses zijn uitgevoerd in SPSS (versie 15.0). Mogelijke problemen van collineariteit zijn onderzocht omdat alle indicatoren werden berekend op dezelfde dataset. Gelet op de veelal scheve verdeling van de scores zijn non-parametrische correlatie coëfficiënten gebruikt (kendall's tau-b). Om ook de correlatie structuur te verkennen wanneer bepaalde afkappunten zouden worden gebruikt, zijn phi-waarden berekend voor scores gesplitst op de mediaan. De correlatie matrix werd ingebracht in een factor analyse waarbij een oblimin rotatie met Kaiser normalisatie werd toegepast om samenhang van de factoren aan te geven (Mittag, 1993).

Diverse classificaties van continuïteitsmaten en kenmerken van zorgpatronen (Jee & Cabana, 2006; Sytma et al., 1989) suggereren een clustering van bepaalde indicatoren. Indexen die concentratie of fragmentatie van contacten over verschillende zorgvoorzieningen uitdrukken, zijn in verband gebracht met respectievelijk informatie en management continuïteit. Relationale continuïteit wordt geoperationaliseerd als opeenvolgende contacten met dezelfde hulpverlener en contact continuïteit wordt gemeten met de duur van de begeleidingsperiode of frequentie van contacten. In tabel 2 zijn de verwachte clusters aangegeven en de factorladingen en communaliteiten van de analyse met vier factoren, die samen een variantie van 72,9% binden. De eigenwaarden en scree plot geven aan dat drie of vier factoren konden worden onderscheiden. Nadere analyses met andere aantallen factoren en het weglaten van

indicatoren met hoge correlaties gaven vergelijkbare resultaten te zien. Sommige communaliteiten waren lager dan .40, wat betekent dat de betreffende indicator weinig met andere maten gemeenschappelijk heeft. Op basis van een drempelwaarde van .40 voor de factorladingen blijkt alleen de duur van de relatie tussen patiënt en hulpverlener (DOR) niet bij een factor te kunnen worden ingedeeld.

In tabel 2 is factor 1 gelabeld als informatiecontinuïteit en deze is gecorreleerd (-.55) met factor 4 waarin managementcontinuïteit wordt uitgedrukt. De samenhang blijkt ook uit de relatief hoge ladingen op beide factoren van de indexen CFC en FRAC. De factoren 2 en 3 zijn geïnterpreteerd als respectievelijk contact en relationele continuïteit. De factorcorrelaties varieerden van .26 en -.36 voor informatie- en managementcontinuïteit met relatiecontinuïteit, tot .01 voor relatie- en contactcontinuïteit. Deze relatief lage factorcorrelaties suggereren dat indicatoren zoals CIA, CONI, FRAC, SECON en FSM of SCi bruikbaar zijn voor het monitoren van verschillende typen van continuïteit van zorg.

#### 4. *Discussie*

Consensus groeit over de multidimensionaliteit van continuïteit van zorg en er zijn bruikbare, objectieve indicatoren voorhanden. De toepassing van dergelijke continuïteitsmaten in kwaliteitsbeoordelingen van de GGz is echter nog beperkt omdat de beschikbare informatie vaak nog veel te wensen overlaat. Rosenbeck et al. (1999) benoemen de bouwstenen, of valkuilen, voor een bruikbare database voor GGz-monitoring. De gegevensverzameling moet longitudinaal zijn en alle belangrijke zorgfuncties omvatten, relevante uitkomstmaten in beeld brengen, standaard meetinstrumenten gebruiken, ook relevante verklarende variabelen vastleggen, en diverse settings of zorgorganisaties omvatten.

Psychiatrische casusregisters zijn in staat om verschillende typen van zorgcontinuïteit *longitudinaal en integraal* te monitoren. Ook geïntegreerde GGz voorzieningen hebben de mogelijkheid om verschillende continuïteitsmaten mee te nemen in de beoordeling van de kwaliteit van de zorg. De technische mogelijkheden om gegevens longitudinaal te verwerken zijn in de afgelopen decennia sterk uitgebreid. Echter, de GGz, en in het bijzonder de openbare GGz, functioneert steeds meer in samenhang met aangrenzende zorgsectoren, zoals de eerstelijnszorg, jeugdzorg en maatschappelijke opvang. Wanneer patiënten niet in relevante zorgsectoren kunnen worden gevolgd omdat dergelijke informatie ontbreekt of koppeling van gegevens niet uitvoerbaar is, zijn verschillen in de tijd en tussen regio's niet goed vergelijkbaar (Wierdsma & Sytema, 1996).

*Relevante uitkomstmaten* moeten het longitudinale karakter van continuïteit van zorg tot uitdrukking brengen. Continuïteitsmaten die het patroon van contacten van een patiënt over een bepaalde observatieperiode tot één cijfer reduceren, roepen de kritiek op dat het beloop van het zorggebruik niet wordt gemeten (Hag-



gerty et al., 2003). Dit beloop kan worden onderzocht met 'time variant measures', waarbij herhaalde metingen worden gecreëerd door per datum in de onderzoeksperiode een score over bijvoorbeeld de voorafgaande maand te berekenen (Fortney et al., 2003). Met de bewerking en analyse van deze metingen is echter nog weinig ervaring opgedaan - zie voor een voorbeeld De Vries en Wierdsma (2009).

De inventarisatie van indicatoren van zorgpatronen laat zien dat de verschillende typen continuïteit kunnen worden uitgedrukt in *standaard maten*. Het berekenen van deze continuïteitsmaten en de analyse van de uitkomsten, in het bijzonder longitudinale vergelijkingen, vereist echter een combinatie van inhoudelijke en technische kennis die niet tot de standaard uitrusting van zorgonderzoekers behoort (Goldman et al., 1999). De financiering van een dergelijke complexe informatievoorziening wordt bovendien voortdurend bedreigd door de korte termijn informatiebehoeften van beleidsmakers (Wierdsma, 1995).

Om verschillen op de uitkomstmaten te kunnen interpreteren, moeten ook *relevante verklarende variabelen* worden geregistreerd of aan de dataset kunnen worden gekoppeld. Routine Outcome Monitoring wordt binnenkort verplicht in de GGz. Daartoe worden herhaalde metingen verricht met standaard meetinstrumenten zoals de Health of the Nation Outcome Scales (HoNOS). Dit is een in Engeland ontwikkelde itemlijst waarmee de gezondheid en het sociaal functioneren van patiënten wordt gevolgd (Mulder et al., 2000; Wierdsma, 1995). Een voorwaarde is dan wel dat zorginstellingen dezelfde methode gebruiken, waartoe recent een aanzet is gegeven in een consensusdocument (Mulder et al., 2010).

*Multisite research* is nodig om de beperkingen te overstijgen van de resultaten van een unieke samenstelling van voorzieningen binnen één regio. De bestaande psychiatrische casusregisters bieden mogelijkheden om regionale verschillen in de GGz en de ontwikkeling van het zorggebruik te monitoren. Er zijn echter weinig redenen voor onderzoekers om zich met deze onderwerpen bezig te houden. De mogelijkheden voor publicatie en het aantrekken van derde geldstroom onderzoek zijn beperkt. Na het succesvolle NWO stimuleringsprogramma 'Geestkracht' wordt het wellicht tijd voor een vervolg met 'Zorgpracht'.

## Referenties

Bachrach, L. (1981). Continuity of care for chronic mental patients: a conceptual analysis. *American Journal of Psychiatry*, 138, 1449-1456.

Bice, T. W., & Boxerman, S. B. (1977). A Quantitative Measure of Continuity of Care. *Medical Care*, 15,4, 347-349.

Blansjaar, B. A. (1992). *Alcoholic Korsakoff's Syndrome*. Rijksuniversiteit Leiden, Leiden.

- Breslau, N., & Reeb, K. (1975). Continuity of care in a university-based practice. *Journal of Medical Education*, 50,10, 965–969.
- Brown, L., Cox, G. B., Jones, W. E., Semke, J., Allen, D. G., Gilchrist, L. D., & Sutphen-Mroz, J. (1994). Effects of mental health reform on client characteristics, continuity of care and community tenure. *Evaluation and Program Planning*, 17,1, 63–72.
- Carr, V. J., Johnston, P. J., Lewin, T. J., Rajkumar, S., Carter, G. L., & Issakidis, C. (2003). Patterns of Service Use Among Persons With Schizophrenia and Other Psychotic Disorders. *Psychiatric Services*, 54,2, 226–235.
- Citro, R., Ghosh, S., & CHurgin, P. (1997). A fundamental metric for continuity of care: modeling and performance evaluation. *IEEE Transactions on Information Technology in Biomedicine*, 1,3, 189–204.
- Dieperink, C. J., Pijl, Y. J., & Driessen, G. (2006). Langdurig zorgafhankelijken in de GGZ. *Maandblad Geestelijke volksgezondheid*, 61,3, 228–238.
- Ejlertsson, G., & Berg, S. (1984). Continuity-of-Care Measures. An Analytic and Empirical Comparison. *Medical Care*, 22,3, 231–239.
- Fortney, J., Sullivan, G., Williams, K., Jackson, C., Morton, S. C., & Koegel, P. (2003). Measuring continuity of care for clients of public mental health systems. *Health Services Research*, 38,4, 1157–1175.
- Freeman, G. K., Weaver, T., Low, J., de Jonge, E., & Crawford, M. (2002). *Promoting Continuity of Care for People with Severe Mental Illness whose needs span primary, secondary and social care. A multi-method investigation of relevant mechanisms and contexts: (NCCSDO)*.
- Gezondheidsraad. (2004). *Noodgedwongen. Zorg voor niet-opgenomen acute psychiatrische patiënten*. Den Haag: Gezondheidsraad.
- Giel, R., & Sturmans, F. (Eds.). (1996). *Psychiatrische Casus-Registers in Nederland*. Groningen: Rijksuniversiteit Groningen.
- Goldman, W., Sturm, R., & McCulloch, J. (1999). New research alliances in the era of managed care. *The Journal of Mental Health Policy and Economics*, 2, 107–110.
- Gulliford, M. C., Naithani, S., & Morgan, M. (2006). Measuring Continuity of Care in Diabetes Mellitus: An Experience-Based Measure. *Annals of Family Medicine*, 4, 548–555.

- Haggerty, J., Reid, R., Freeman, G., Starfield, B., Adair, C., & McKendry, R. (2003). Continuity of care: a multidisciplinary review. *British Medical Journal*, 327, 1219–1221.
- Hall, S. M., Tunis, S., Triffleman, E., Banys, P., Clark, W., Tusel, D., Stewart, P., & Presti, D. (1994). Continuity of care and Desipramine in primary cocaine abusers. *Journal of Nervous and Mental Disease*, 182, 570–575.
- Hussein, A., & Carrière, K. C. (2002). A measure of continuity of care based on the multiplicative intensity model. *Statistics in Medicine*, 21, 457–465.
- Jee, S. H., & Cabana, M. D. (2006). Indices for Continuity of Care: A Systematic Review of the Literature. *Medical Care Research and Review*, 63,2, 158–188.
- Lavik, N. J. (1983). Utilisation of mental health services over a given period. *Acta Psychiatrica Scandinavica*, 67, 404–413.
- Lehman, A., Postrado, L., Roth, D., McNary, S., & Goldman, H. (1994 ). *Continuity of Care and Client Outcomes in the Robert Wood Johnson Foundation Program on Chronic Mental Illness*. The Milbank Quarterly.
- Lou, W. Y. W. (2000). A New Measure for Continuity of Care: The Alpha Index. *Health Services & Outcomes Research Methodology*, 1,3–4, 277–289.
- Magill, M. K., & Senf, J. (1987). A new method for measuring continuity of care in family practice residencies. *Journal of Family Practice*, 24,2, 165–168.
- McWhinney, I. R., Bass, M. J., Donner, A., & Hoddinott, S. (1988). Is population mobility an obstacle to continuity of care? Attrition rates over five years in 17 Ontario practices. *Journal of Family Practice*, 27,3, 291–296.
- Mittag, K. C. (1993). *Scale-Free Nonparametric Factor Analysis: A User-Friendly Introduction with Concrete Heuristic Examples*. Paper presented at the Annual meeting of the Southwest Educational Research Association.
- Moreno, B., Cervilla, J., Luna, J. D., & Torres, F. (2007). Pattern of care for schizophrenia patients in Granada (Spain): a case register study. *Int J Soc Psychiatry*, 53,1, 5–11.
- Mulder, C., Koopmans, G., & Selten, J. (2006). Emergency psychiatry, compulsory admissions and clinical presentation among immigrants to the Netherlands. *British Journal of Psychiatry* Apr; 188, 386–391.

Mulder, C. L., Sytema, S., & Wierdsma, A. I. (2000). Statusmeting en instrument-gestuurde planning in de GGZ. Een zoektocht naar de heilige graal. *Maandblad Geestelijke volksgezondheid*, 55,9, 790–799.

Mulder, C. L., van der Gaag, M., Bruggeman, R., Cahn, W., Delespaul, P. A. E., Dries, P., Faber, G., de Haan, L., van der Heijden, F. M. M. A., Kempen, R. W., Moggendorff, E. S. G., Slooff, C. J., Sytema, S., Wiersma, D., Wunderink, L., & van Os, J. (2010). Routine outcome monitoring voor patiënten met ernstige psychiatrische aandoeningen; een consensusdocument. *Tijdschrift voor Psychiatrie*, 52,3, 169–179.

Pijl, Y. J., Driessen, G. A. M., Wierdsma, A. I., & Sytema, S. (2005). *Evaluatie van het extramuraliseringsbeleid in de GGZ*. Utrecht: Trimbos-instituut.

Reid, R., Haggerty, J., & McKendry, R. (2002). *Defusing the Confusion: Concepts and Measures of Continuity of Healthcare*: Canadian Health Services Research Foundation.

Rigter, H., Have, M. t., Cuijpers, P., Depla, M., Gageldonk, A. v., Laan, G. v. d., Peterse, A., Ruiter, C. d., Smits, C., & Wolf, J. (2002). *Brancherapport GGZ-MZ 1998–2001*. Den Haag: Sdu Uitgevers.

Roos, L. L., Roos, N. P., Gilbert, P., & Nicol, J. P. (1980). Continuity of Care: Does It Contribute to Quality of Care? *Medical Care*, 18,2, 174–184.

Rosenheck, R., Fontana, A., & Stolar, M. (1999). Assessing quality of care: Administrative indicators and clinical outcomes in posttraumatic stress disorder. *Medical Care*, 37,2, 180–188.

Schnabel, P. (1992). Wonen en werken met Korsakov. In Verslag van het seminar "Wonen en werken met Korsakov". Rotterdam: SOGG.

Schnabel, P. (1995). *De weerbarstige geestesziekte, naar een nieuwe sociologie van de geestelijke gezondheidszorg*. Nijmegen: SUN.

Scholten, C. M., & Tjadens, F. L. J. (1996). *Wet BOPZ: Verkommerden en verloederden*. Rijswijk: Ministerie van VWS.

Schrag, D., Xu, F., Hanger, M., Elkin, E., Bickell, N. A., & Bach, P. B. (2006). Fragmentation of Care for Frequently Hospitalized Urban Residents. *Medical Care*, 44,6, 560–567.

Selten, J.-P., Wierdsma, A., Mulder, N., & Burger, H. (2007). Seeking treatment for alcohol and drug-use disorders by immigrants to the Netherlands. Retrospective, population-based, cohort study. *Social Psychiatry and Psychiatric Epidemiology*, 42, 301–306.

Shortell, S. A. (1976). Continuity of Medical Care: Conceptualization and Measurement. *Medical Care*, 14, 377–391.

Singh, S. P., Greenwood, N., White, S., & Churchill, R. (2007). Ethnicity and the Mental Health Act 1983. *British Journal of Psychiatry*, 191, 99–105.

Spooner, S. A. (1994). Incorporating temporal and clinical reasoning in a new measure of continuity of care. Paper presented at the *Proceedings of the Annual Symposium on Computer Application in Medical Care*.

Steinwachs, D. M. (1979). Measuring provider continuity in ambulatory care. *Medical Care*, 17, 551–565.

Stobbe, J., Wierdsma, A. I., van Beest, R. H. P., & Mulder, C. L. (2009). Drop-out na gedwongen opname – hoe groot is het probleem? *Tijdschrift voor Psychiatrie*, 51,11, 801–812.

Sytema, S., Brook, F. G., Giel, R., & Ten Horn, G. H. M. M. (1985). Patiënten met een neurotische depressie in de GGZ. *Tijdschrift voor Psychiatrie*, 27, 237–252.

Sytema, S., Burgess, P., & Tansella, M. (2002). Does community care decrease length of stay and risk of rehospitalization in new patients with schizophrenic disorders? A comparative Groningen (The Netherlands), Victoria (Australia) and Verona (Italy) case-register study. *Schizophrenia Bulletin*, 28,2, 273–281.

Sytema, S., Giel, R., & Ten Horn, G. H. M. M. (1989). Patterns of care in the field of mental health. Conceptual definition and research methods. *Acta Psychiatrica Scandinavica*, 79,1, 1–10.

ten Horn, G. H. M. M. (1982). Nazorg geeft kopzorg. Rijksuniversiteit Groningen, Groningen.

ten Horn, G. H. M. M. (1989). The development of other European case registers. In J. K. Wing (Ed.), *Health Services Planning and Research. Contributions from Psychiatric Case Registers*. London: Gaskell.

ten Horn, G. H. M. M., & Haffner, J. (1989). *The types of treatment and care psychiatric patients receive: an attempt to standardize different measures and patterns of care*. Groningen.

Vermande, M. M., & Bijl, R. V. (1995). *Inventarisatie Regionale Informatiesystemen Patiëntgegevens GGz*. Utrecht: Nederlands centrum Geestelijke volksgezondheid.

Vries, S. C. d., & Wierdsma, A. I. (2009). Exploring Continuity of Care in Patients with Alcohol Use Disorders Using Time-Variant Measures. *European Addiction Research*, 15, 19–24.

Wierdsma, A. I. (1995). Psychiatrische casusregisters in Engeland: Toen en nu. *Tijdschrift Sociale Gezondheid*, 73,7, 469–473.

Wierdsma, A. I., Mulder, C. L., de Vries, S., & Sytema, S. (2009). Reconstructing continuity of care in mental health services: a multilevel conceptual framework. *Journal of Health Services Research & Policy*, 14,1, 52–57.

Wierdsma, A. I., Pijl, Y. J., Driessen, G. A. M., & Sytema, S. (2006). Extramuralisering van de GGZ in Nederland: landelijk beleid of regiovisie? *Maandblad Geestelijke volksgezondheid*, 61,5, 427–434.

Wierdsma, A. I., & Sytema, S. (1996). Regio's vergelijkenderwijs: waarop moet worden gelet? In R. Giel & F. Sturmans (Eds.), *Psychiatrische Casus-Registers in Nederland* (pp. 41–53): RU Groningen.

Wierdsma, A. I., Sytema, S., van Os, J. J., & Mulder, C. L. (2007). Hebben we psychiatrische casusregisters nog nodig? *Tijdschrift voor Psychiatrie*.

Wierdsma, A. I., van Baars, A. W. B., & Mulder, C. L. (2006). Psychiatrische voorgeschiedenis en nazorg bij dwangopneming. Zorggebruik als indicator van de kwaliteit van zorg bij inbewaringstellingen in Rotterdam. *Tijdschrift voor Psychiatrie*, 48,2, 81–93.

Wierdsma, A. I., van Marle, P. D., & Mulder, C. L. (2010). Aantallen en patronen van inbewaringstellingen en rechterlijke machtigingen in Maastricht, Groningen en Rotterdam. Bijdrage van stedelijke problematiek aan regionale verschillen in Bopz-maatregelen. *Tijdschrift voor Psychiatrie*, 52,3, 143–153.

Wolinsky, F. D., Miller, T. R., Geweke, J. F., Chrischilles, E. A., An, H., Wallace, R. B., Pavlik, C. E., Wright, K. B., Ohsfeldt, R. L., & Rosenthal, G. E. (2007). An Interpersonal Continuity of Care Measure for Medicare Part B Claims Analyses. *Journal of Gerontology*, 62B,3, S160–S168.

# *Het meten van de kwaliteit van administratieve bronnen: Recente resultaten en toekomstige ontwikkelingen*

Piet J.H. Daas, Saskia J.L. Ossen en Martijn Tennekes (CBS)

## *1. Inleiding*

Nationale Statistische Instituten (NSI's) hebben gegevens nodig om statistieken te kunnen maken. Veel van die gegevens worden met behulp van vragenlijsten verzameld. Steeds vaker maken NSI's echter ook gebruik van gegevens uit bronnen die door andere organisaties verzameld zijn. Voorbeelden van dergelijke bronnen zijn registers en administratieve bronnen (Wallgren en Wallgren, 2007). De gegevens in die bronnen worden gebruikt voor en zijn het gevolg van administratieve processen. In de praktijk blijken ze vaak ook erg interessant te zijn voor de statistiek. Dit besef is in de afgelopen 10 jaar bij steeds meer NSI's doorgedrongen (Unece 2007). De belangrijkste voordelen van het gebruik van administratieve bronnen en registers voor de statistiek zijn: i) reductie van de kosten van het verzamelen van de gegevens en ii) lastendrukvermindering voor bedrijven en personen. Omdat administratieve bronnen vaak gegevens over hele populaties bevatten, in verschillende tijdsperiodes, zijn ze ook erg geschikt om te gebruiken voor (virtuele) volkstellingen (Schulte Nordholt, 2004), voor gedetailleerde longitudinale analyse van (sub)populaties en regio's (Wallgren en Wallgren, 2007) en voor cohortonderzoek van studenten (Chowdry et al., 2010).

Vanuit statistisch oogpunt bekeken kleven aan het gebruik van administratieve bronnen echter ook enkele nadelen. Deze zijn voornamelijk het gevolg van het feit dat de verzameling en verwerking van de gegevens niet door het NSI wordt uitgevoerd. Het is de beheerder van de bron (de 'bronhouder') die deze taken verricht. Een soortgelijk probleem treedt op bij de eenheden en variabelen die in een administratieve bron zijn opgeslagen. De definitie hiervan wordt uitsluitend door de administratieve regels van de bronhouder bepaald. Hierdoor kunnen de gehanteerde definities afwijken van degene die door het NSI worden gebruikt (Wallgren en Wallgren, 2007). Het is dan ook niet altijd even eenvoudig om de statistische bruikbaarheid van een administratieve bron te bepalen (Bakker et al., 2008).

Omdat de kwaliteit van statistieken sterk wordt beïnvloed door de kwaliteit van de gegevens die aan het begin van het statistisch proces liggen, is het van essentieel

belang dat NSI's de kwaliteit van administratieve bronnen eenduidig en efficiënt kunnen bepalen. Deze constatering vormde de aanleiding voor het ontwikkelen van een kwaliteitsraamwerk voor administratieve bronnen en registers op het Centraal Bureau voor de Statistiek (CBS). Dit raamwerk moet het mogelijk maken om de statistische bruikbaarheid (de kwaliteit) van extern verzamelde, secundaire, gegevens, aan het begin van het proces, op een efficiënte en transparante manier te bepalen (Daas et al., 2008).

## 2. *Kwaliteitsraamwerk*

Op het CBS is een uitgebreide literatuurstudie uitgevoerd om de verschillende kwaliteitsaspecten van administratieve bronnen te identificeren. Deze studie liet zien dat het perspectief op de kwaliteit van dergelijke bronnen in de diverse publicaties erg kan verschillen (Daas et al., 2008). Afhankelijk van het gehanteerde perspectief traden andere kwaliteitsaspecten op de voorgrond. Deze constatering is op zich niet nieuw. Dergelijke perspectieven worden vaak categorieën (Batini en Scanapiego, 2006) of hyperdimensies genoemd (Karr et al., 2006). De laatste term zal in de rest van dit hoofdstuk worden gebruikt.

De verschillende perspectieven die in de literatuurstudie werden geïdentificeerd bleken met drie hyperdimensies volledig beschreven te kunnen worden (Daas et al., 2008). De hyperdimensies werden Bron, Metadata en Data genoemd. De drie hyperdimensies vormen de basis van het ontwikkelde kwaliteitsraamwerk voor administratieve bronnen en registers. Elke hyperdimensie in het raamwerk is opgebouwd uit een aantal dimensies, waarbij elke dimensie een aantal kwaliteitsindicatoren bevat. Een kwaliteitsindicator wordt gemeten of geschat door één of meerdere meetmethoden die kwantitatief of kwalitatief kunnen zijn (Daas et al., 2008). Dit hoofdstuk begint met een bespreking van de kwaliteitsaspecten en de daarbij behorende meetmethoden die voor de hyperdimensies Bron en Metadata zijn ontwikkeld. Vervolgens worden inzichten beschreven voor het bepalen van de kwaliteit van de gegevens die tot de hyperdimensie Data behoren.

### 2.1 *Hyperdimensies Bron en Metadata*

Een NSI dat van plan is een administratieve bron als inputbron voor de statistiek te gaan gebruiken dient als eerste de kwaliteitsaspecten die met de levering van de bron te maken hebben te bepalen. Deze kwaliteitsaspecten behoren tot de Bron hyperdimensie van het kwaliteitsraamwerk. Tabel 1 geeft een overzicht van de dimensies, kwaliteitsindicatoren en meetmethoden voor de Bron hyperdimensie.

De Metadata hyperdimensie richt zich op de conceptuele en proces gerelateerde kwaliteitsaspecten van de metadata van de bron. Het is uitermate belangrijk dat een NSI de metadata gerelateerde kwaliteitsaspecten volledig begrijpt. Elk misver-



**Tabel 1**  
**Kwaliteitsraamwerk voor registers, hyperdimensie Bron**

DIMENSIES	KWALITEITSINDICATOREN	MEETMETHODEN
1. Leverancier	1.1 Contact	<ul style="list-style-type: none"> <li>- Naam databron</li> <li>- Contactgegevens bronhouder/beheerder</li> <li>- Contactpersoon NSI</li> </ul>
	1.2 Doel	<ul style="list-style-type: none"> <li>- Reden gebruik bron door NSI</li> </ul>
2. Relevantie	2.1 Nut	<ul style="list-style-type: none"> <li>- Belang bron voor NSI</li> </ul>
	2.2 Mogelijk gebruik	<ul style="list-style-type: none"> <li>- Potentieel gebruik bron voor statistiek</li> </ul>
	2.3 Informatiebehoefte	<ul style="list-style-type: none"> <li>- Voldoet de bron aan informatiebehoefte v/h NSI</li> </ul>
	2.4 Lastendruk	<ul style="list-style-type: none"> <li>- Gevolgen gebruik op lastendruk van NSI</li> </ul>
3. Privacy en beveiliging	3.1 Wettelijke basis	<ul style="list-style-type: none"> <li>- Grondslag voor bestaan v/d bron</li> </ul>
	3.2 Vertrouwelijkheid	<ul style="list-style-type: none"> <li>- Is WBP van toepassing?</li> <li>- Is gebruik door NSI aangemeld?</li> </ul>
	3.3 Beveiliging	<ul style="list-style-type: none"> <li>- Wijze versturen gegevens naar NSI</li> <li>- Noodzaak beveiliging (soft- en hardware)</li> </ul>
4. Levering	4.1 Kosten	<ul style="list-style-type: none"> <li>- Kosten verbonden aan gebruik door NSI</li> </ul>
	4.2 Afspraken	<ul style="list-style-type: none"> <li>- Is er een leveringsovereenkomst?</li> <li>- Frequentie leveringen</li> </ul>
	4.3 Stiptheid	<ul style="list-style-type: none"> <li>- Hoe stipt kan geleverd worden?</li> <li>- Snelheid doorgifte afwijkingen</li> <li>- Snelheid opslag gegevens bij bronhouder</li> </ul>
	4.4 Opmaak	<ul style="list-style-type: none"> <li>- Format(s) waarin data geleverd kan worden</li> </ul>
	4.5 Selectie	<ul style="list-style-type: none"> <li>- Welke gegevens kunnen geleverd worden?</li> <li>- Is dit wat het NSI wil hebben?</li> </ul>
5. Procedures	5.1 Data verzamelen	<ul style="list-style-type: none"> <li>- Bekendheid met wijze van data verzamelen</li> </ul>
	5.2 Wijzigingsplannen	<ul style="list-style-type: none"> <li>- Bekendheid met wijzigingsplannen</li> <li>- Wijze van communiceren met NSI</li> </ul>
	5.3 Terugkoppeling	<ul style="list-style-type: none"> <li>- Mag NSI bij problemen terugkoppelen?</li> <li>- Wat wel, wat niet en waarom?</li> </ul>
	5.4 Terugvalscenario	<ul style="list-style-type: none"> <li>- Afhankelijkheidsrisico v/h NSI</li> <li>- Maatregelen bij het niet leveren volgens afspraak</li> </ul>

stand of fout hierin zal de kwaliteit van de geproduceerde statistieken aanzienlijk beïnvloeden. In tabel 2 zijn de dimensies, kwaliteitsindicatoren en meetmethoden van de Metadata hyperdimensie weergegeven.

## 2.2 Checklist voor Bron en Metadata

Voor de verschillende kwaliteitsaspecten in de hyperdimensies Bron en Metadata is een checklist ontwikkeld (Daas et al., 2009a). De Engelstalige versie van de checklist is in de CBS-publicatie van Daas et al. (2009b) opgenomen; deze publicatie is te vinden op de website van het CBS in de rubriek 'discussion papers'. De checklist leidt de gebruiker door de meetmethoden van de kwaliteitsindicatoren in de hyperdimensies Bron en Metadata. Door de vragen in de checklist, voor de betreffende verslagperiode, te beantwoorden wordt de 'waarde' voor elke meetmethode in tabel 1 en 2 bepaald.

Bij de evaluatie van het Metadata-deel in de checklist is het noodzakelijk dat de gebruiker een specifieke statistiek waarvoor de bron wordt gebruikt in gedachten heeft. Dit is nodig omdat in dit deel de definities van de eenheden, variabelen en tijdsperiode(n) van de bronhouder vergeleken worden met de definities zoals ze door de betreffende statistiek worden gebruikt.

**Tabel 2**  
**Kwaliteitsraamwerk voor registers, hyperdimensie Metadata**

DIMENSIES	KWALITEITSINDICATOREN	MEETMETHODEN
1. Duidelijkheid	1.1 Populatie definitie	- Score duidelijkheid omschrijving
	1.2 Classificatievariabele definitie	- Score duidelijkheid omschrijving
	1.3 Telvariabele definitie	- Score duidelijkheid omschrijving
	1.4 Tijdsdimensie definitie	- Score duidelijkheid omschrijving
	1.5 Definitiewijzigingen	- Bekendheid met opgetreden wijzigingen
2. Vergelijkbaarheid	2.1 Populatie definitie	- Vergelijkbaarheid met NSI-definitie
	2.2 Classificatievariabele definitie	- Vergelijkbaarheid met NSI-definitie
	2.3 Telvariabele definitie	- Vergelijkbaarheid met NSI-definitie
	2.4 Tijdsverschillen	- Vergelijkbaarheid met tijdsperioden NSI
3. Unieke sleutels	3.1 Identificerende sleutels	- Aanwezigheid unieke sleutels - Overeenkomst met unieke sleutels van NSI
	3.2 Unieke combinaties	- Aanwezigheid bruikbare combinaties van variabelen
4. Databehandeling (door bronhouder)	4.1 Controles	- Gebruikte controles van populatie eenheden - Gebruikte controle van variabelen - Controles op combinaties van variabelen - Controles op extreme waarden (uitbijters)
	4.2 Aanpassingen/bewerkingen	- Bekendheid met aanpassingen/bewerkingen - Worden aangepaste velden gemarkeerd? - Bekendheid met gebruik van standaardwaarden

Om de bruikbaarheid van de checklist te testen zijn acht secundaire databronnen van het CBS geëvalueerd. Deze bronnen zijn: de Polisadministratie (PA), het bestand Wet op de Studiefinanciering (WSF), de gegevens van werkzoekenden van het Centrum voor Werk en Inkomen (CWI; tegenwoordig het UWV WERKbedrijf), het Examen Resultatenregister (ERR), het gecoördineerde eencijferregister afgeleid uit het Centraal Register Inschrijvingen Hoger Onderwijs (1CijferHO), het gecoördineerde eencijferregister afgeleid uit de onderwijsnummerbestanden voor het Voortgezet Onderwijs (1CijferVO), de Nationale Autopas gegevens (NAP) en de Gemeentelijke Basisadministratie persoonsgegevens (GBA).

Omdat het belangrijkste doel van de studie de bruikbaarheid van de resultaten van de checklist was, werden de vragenlijsten door de gebruikers ingevuld in samenwerking met één of meerdere auteurs van dit rapport. De gebruikers waren CBS-medewerkers die betrokken zijn bij: contact met de bronhouder, ontvangst van de databron, en/of verwerking en controle van de bron. Gemiddeld duurde het ongeveer 2 uur om de checklist voor een bron in te vullen.

### 2.3 Resultaten checklist

De resultaten voor de acht databronnen zijn in tabellen 3 en 4 weergegeven. Tabel 3 bevat de resultaten voor de Bron hyperdimensie en tabel 4 die voor de Metadata hyperdimensie. Voor de PA is het Metadata-deel van de checklist ingevuld met het oog op gebruik voor de werkloosheidsstatistiek. De GBA is beoordeeld met gebruik voor de bevolkingsstatistiek in gedachten, terwijl de NAP bekeken is met beoogd gebruik voor de verkeer- en vervoersstatistiek. De andere bronnen zijn beoordeeld met het oog op gebruik voor het schatten van het opleidingsniveau van de Nederlandse bevolking (Bakker et al., 2008)

**Tabel 3**  
Resultaten voor de hyperdimensie Bron

	Databronnen							
	PA	WSF	CWI	ERR	1FigHO	1FigVO	NAP	GBA
<i>Dimensies</i>								
1. Leverancier	+	+	+	+	+	+	+	+
2. Relevantie	+	+	+	o	+	+	+	+
3. Privacy en beveiliging	+	+	+	+	+	+/o	+	+
4. Levering	o	+	-	+	+	o	+	+
5. Procedures	+	+/o	+	+/o	+/o	+/o	o	+

**Tabel 4**  
Resultaten voor de hyperdimensie Metadata

	Databronnen							
	PA	WSF	CWI	ERR	1FigHO	1FigVO	NAP	GBA
<i>Dimensies</i>								
1. Duidelijkheid	+	+	-	o	+	+	+	+
2. Vergelijkbaarheid	+/o	+	-	+	+	+	+	+
3. Unieke sleutels	+	+	+	+	+	+	+	+
4. Data behandeling	+/o	?(+)	?	?(o)	?(+)	?(+)	+	+

De resultaten van de evaluatie zijn in tabel 3 en 4 op het niveau van de dimensies weergegeven. De getoonde scores werden bepaald door de meest voorkomende score van de meetmethoden die tot die dimensie behoren te selecteren. De gebruikte symbolen voor de scores in tabel 3 en 4 zijn: goed (+), redelijk (o), slecht (-) en onduidelijk (?). Tussensliggende scores worden weergegeven door de symbolen te combineren met een scheidingsteken (/). Wanneer een onduidelijk score in een bepaalde dimensie voorkwam is dit resultaat voor de gehele dimensie getoond. Echter, wanneer daarnaast bleek dat alle andere meetmethoden in die dimensie wel duidelijk scoorden is de meest voorkomende score van die andere methoden tussen haakjes toegevoegd.

### *Hyperdimensie Bron*

De resultaten in tabel 3 laten zien dat de scores van alle databronnen, voor het Bron-deel van de checklist, ietwat laag zijn voor de dimensies Levering en Procedures. Voor de dimensie Levering is dit voornamelijk het gevolg van het niet altijd tijdig leveren van de PA, CWI en 1FigHO bestanden. Dit duidt op een mogelijk risico voor CBS-gebruikers die erg afhankelijk zijn van de tijdige levering van deze bronnen. Het grootste probleem in de hyperdimensie Bron vormt de tijdige beschikbaarheid van de CWI. Deze bron wordt vrijwel nooit op tijd geleverd; een vertraging van enkele uren, dagen of zelfs weken is eerder regel dan uitzondering. Er is zelfs eens een periode van 3 maanden geweest waarin door de beheerder geen gegevens werden geleverd. De ietwat lage scores in de dimensie Levering voor de

PA en 1FigHO zijn overigens niet geheel onverwacht. Beide bronnen zijn nog niet helemaal uitontwikkeld; ze bestaan nog niet zo heel lang. Hierdoor fluctueren de leveringstijden nog enigszins.

In de dimensie Procedures zijn de scores ietwat laag door de relatief lage score op de terugvalscenario indicator (indicator nr. 5.4 in tabel 1). Niet alle CBS-gebruikers waren met de CBS-regel bekend dat een dergelijk scenario niet voor alle administratieve bronnen hoeft te worden opgesteld. Dit beïnvloedde de score voor de dimensie Procedures in negatieve zin. Wanneer hiermee rekening wordt gehouden komen er in deze dimensie, op één uitzondering na, eigenlijk nauwelijks problemen voor. De uitzondering was de NAP, hierbij verloopt het contact met de bronhouder wat moeizaam. Verzoeken om extra informatie worden niet altijd tijdig beantwoordt en de antwoorden die worden gegeven zijn niet altijd even verhelderend.

### *Hyperdimensie Metadata*

De resultaten voor het Metadata-deel van de checklist zijn in tabel 4 weergegeven. Vergeleken met de resultaten voor Bron (tabel 3) zijn er duidelijk meer slechte (-) scores te zien. Deze scores hebben ook hier weer betrekking op het CWI. Deze databon scoort slecht in de dimensies Duidelijkheid en Vergelijkbaarheid. Dit is voor beide dimensies voornamelijk het gevolg van een verschil tussen de definitie van de CWI-variabele 'opleidingsniveau' en de corresponderende CBS-variabele. Bij het CWI is deze variabele namelijk duidelijk minder strikt gedefinieerd dan bij het CBS. Bij het CWI gaat het bij deze variabele ruwweg om het best bemiddelbare opleidingsniveau. Stel bijvoorbeeld dat iemand op een universiteit is afgestudeerd in een op dat moment lastig bemiddelbare studierichting. Dan kan het voorkomen dat deze persoon omgeschoold wordt, waarbij de omscholing gebeurt op een lager niveau dan universitair. Op het moment dat de omscholing voltooid is, zal het CWI als opleidingsniveau het niveau van de omscholing vermelden. Dit opleidingsniveau wordt dan immers beter bemiddelbaar geacht. In dat geval registreert het CWI dus een lager niveau dan het hoogstbehaalde (Bakker et al., 2008). Ook komt het voor dat het opleidingsniveau dat bij het CWI vermeld is, hoger is dan het hoogst behaalde opleidingsniveau. De geregistreerde in kwestie heeft dan slechts onderwijs genoten op een bepaald opleidingsniveau, maar de betreffende opleiding nooit met een diploma afgesloten.

Uit het voorgaande verhaal blijkt dat de exacte definitie van de variabele opleidingsniveau van het CWI ondermeer afhankelijk is van de arbeidsmarkt en daarmee dus tijdsafhankelijk is. De gehanteerde relatief losse definitie van opleidingsniveau levert ook problemen op bij het communiceren van een definitiewijziging, omdat het CWI-hoofdkantoor een dergelijke wijziging vaak niet precies kan duiden. Dit maakt dat de CWI ook in de dimensie Vergelijkbaarheid slechts scoort.

Van alle dimensies die tot de Metadata hyperdimensie behoren is de dimensie 'Data behandeling' het meest onduidelijk. Dit geeft aan dat op het CBS relatief weinig bekend is over de mogelijke controles, aanpassingen en bewerkingen die

door de bronhouder worden uitgevoerd. Positieve uitzonderingen hierop zijn de PA, NAP en GBA. Bij deze bronnen is de verkregen kennis over 'data behandeling' echter voornamelijk door praktijkervaring geleerd en niet door gerichte studie. Praktijkervaring is zeer belangrijk omdat er sprake kan zijn van een verschil tussen het protocol en de daadwerkelijke uitvoering bij de bronhouder.

De scores voor de databronnen in tabel 3 en 4 laten zien dat aandacht moet worden geschonken aan het CWI en dan vooral aan de kwaliteitsaspecten die tot de dimensies Levering, Duidelijkheid en Vergelijkbaarheid van die bron behoren. De geconstateerde problemen dienen voor het CWI eerst te worden opgelost voordat enige tijd mag worden besteed aan data gerelateerde kwaliteitsstudies. Pas op het moment dat de Bron- en Metadata-gerelateerde problemen voor het CWI zijn opgelost, heeft het nut om (meer) tijd te besteden aan de bepaling van de kwaliteit van de CWI-gegevens. De positieve resultaten voor de GBA laten zien dat het mogelijk is om elk kwaliteitsaspect in de Bron en Metadata hyperdimensie onder controle te hebben. Voor de andere databronnen kan geconstateerd worden dat sommige kwaliteitsgebieden meer aandacht verdienen, maar dat hierbij in het algemeen geen grote problemen werden gevonden. Voor alle bronnen, met uitzondering van het CWI, is de volgende logische stap dan ook het bepalen van de kwaliteit van de gegevens in de bron. Dit is het onderwerp van de volgende paragraaf.

#### 2.4 *Hyperdimensie Data*

Tabel 5 bevat een lijst van kwaliteitsaspecten die tot de hyperdimensie Data worden gerekend. Ze zijn onder andere het resultaat van de eerder vermelde literatuurstudie (Daas et al., 2008). Veel van de indicatoren in tabel 5 zullen statistici bekend voorkomen, maar sommige mogelijk iets minder. De minder bekende indicatoren worden dan ook kort besproken.

Een aanzienlijk deel van de indicatoren in tabel 5 is gebaseerd op de zogenaamde Representativiteitsindex (R-index). Deze indicator is op het CBS ontwikkeld (Schouten et al., 2009). R-indices meten de mate waarin de samenstelling van de eenheden in een bron, op een bepaald moment, afwijkt van de beoogde doelpopulatie. Bij steekproefonderzoeken is dit een bekend concept. Voor dergelijke onderzoeken betekent representativiteit dat alle eenheden in de doelpopulatie dezelfde kans op responderen hebben. Representativiteit is echter ook bij administratieve bronnen belangrijk, zeker wanneer de samenstelling van de populatie in een dergelijke bron tijdsafhankelijk is. Een voorbeeld hiervan is de samenstelling van de bedrijven die Belasting Toegevoegde Waarde (BTW) gegevens naar de Belastingdienst opsturen. In deze bron varieert de samenstelling van bedrijven gedurende de maandelijkse periode van verzamelen (Ouweland et al., 2009). Dit beïnvloedt de kwaliteit van de data in die bron aanzienlijk. Omdat tijdgerelateerde (kwaliteits) aspecten in R-indices worden meegenomen, is tijdigheid niet als aparte dimensie in tabel 5 vermeld. Ook de dimensie Precisie in tabel 5 wordt door tijdsafhankelijke veranderingen op de populatiesamenstelling beïnvloedt.

**Tabel 5**  
**Kwaliteitsraamwerk voor registers, hyperdimensie Data**

DIMENSIES	KWALITEITSINDICATOREN	METHODE BESCHRIJVING
1. Technische checks	1.1. Leesbaarheid 1.2 Voldoen aan metadata	– Is alle data in de bron toegankelijk? – Voldoet alle data aan de metadata-definitie? – Indien niet, meld de verschillen
2. Overdekking	2.1 Niet populatie eenheden	– Percentage eenheden dat niet tot de doelpopulatie behoort
3. Onderdekking	3.1 Ontbrekende eenheden 3.2 Selectiviteit 3.3 Effect op gemiddelde	– Percentage t.o.v. de doelpopulatie ontbrekende eenheden – R-index voor de samenstelling van de eenheden – Maximale vertekening v/h gemiddelde voor kernvariabelen – Maximale RMSE v/h gemiddelde voor kernvariabelen
4. Koppelbaarheid	4.1 Koppelbare eenheden 4.2 Miskoppelingen 4.3 Selectiviteit 4.4 Effect op gemiddelde	– Percentage eenduidig gekoppelde eenheden – Percentage niet-correct gekoppelde eenheden – R-index voor gekoppelde eenheden – Maximale vertekening v/h gemiddelde voor kernvariabelen – Maximale RMSE v/h gemiddelde voor kernvariabelen
5. Unit-nonrespons	5.1 Eenheden zonder data 5.2 Selectiviteit 5.3 Effect op gemiddelde	– Percentage eenheden waar alle gegevens ontbreken – R-index voor samenstelling v/d eenheden – Maximale vertekening v/h gemiddelde voor kernvariabelen – Maximale RMSE v/h gemiddelde voor kernvariabelen
6. Item-nonrespons	6.1 Ontbrekende velden 6.2 Selectiviteit 6.3 Effect op gemiddelde	– Percentage velden met ontbrekende waarden – R-index voor variabele samenstelling – Maximale vertekening v/h gemiddelde voor variabele – Maximale RMSE v/h gemiddelde voor variabele
7 Meting	7.1 Externe controle 7.2 Onverenigbare records 7.3 Meetfout	– Is een audit of parallele toets uitgevoerd? – Is de inputprocedure getest? – Fractie geschonden edit/controleregels – Omvang van de relatieve meetfout
8. Verwerking	8.1 Gaafmaken 8.2 Imputatie 8.3 Uitbijters	– Fractie herziene/gaafgemaakte velden – Fractie geïmputeerde velden – Fractie velden met uitbijtercorrectie
9. Precisie	9.1 Standaardfout	– Schatter voor standaardfout v/h gemiddelde
10. Gevoeligheid	10.1 Ontbrekende velden 10.2 Selectiviteit 10.3 Totaal effect	– Totaal percentage lege velden – R-index voor samenstelling van totalen – Totale maximale vertekening v/h gemiddelde – Totale maximale RMSE v/h gemiddelde

R-index: Representativiteits index, een indicator die de selectiviteit van de ontbrekende gegevens schat door gebruik te maken van informatie afkomstig uit andere bronnen (Schouten et al., 2009).

RMSE: root mean square error; een veel gebruikte maat in de statistiek om de kwaliteit van een schatter te bepalen. De RMSE is gelijk aan de wortel van de som van de vertekening en de variantie van de schatter.

Een ander belangrijk aandachtspunt voor de Data hyperdimensie is het verschil tussen de input- en outputkwaliteit van de gegevens in secundaire bronnen. Dit verschil treedt uitsluitend op wanneer er sprake is van secundair gebruik van de gegevens in een bron. In een dergelijk geval worden de gegevens van een bron namelijk voor een heel ander doel gebruikt dan waar ze oorspronkelijk voor verzameld zijn. Hierdoor is de oorspronkelijke kwaliteit van de gegevens in de bron (= de kwaliteit van de output volgens de bronhouder) immers, per definitie, niet gelijk aan die van het nieuwe beoogde gebruik (= de kwaliteit van de input volgens het CBS). Wanneer het oorspronkelijke doel en het beoogde doel van gebruik gelijk zijn valt dit verschil weg. Daarnaast is er bij het secundair gebruik van administratieve brongegevens ook nog sprake van kwaliteit van de output. Dit is namelijk de

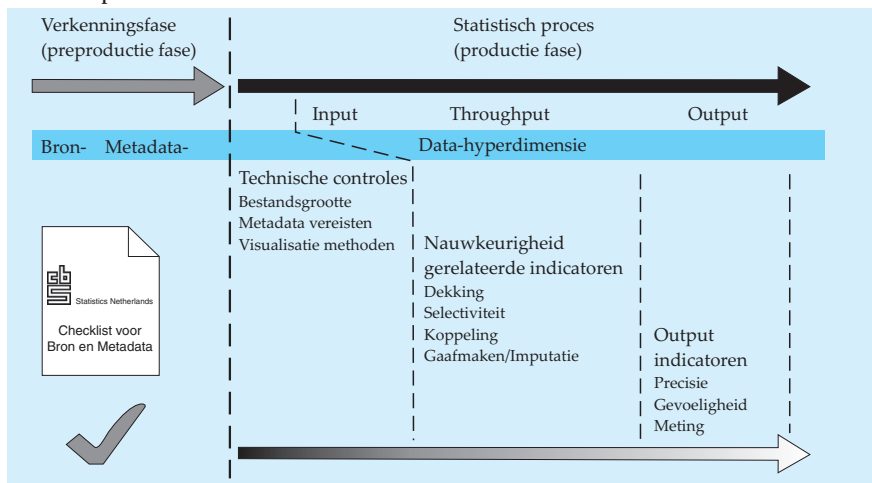
kwaliteit van de statistiek die m.b.v. de secundaire gegevens is gemaakt. Het moge hierbij duidelijk zijn dat de kwaliteit van de statistische output door de kwaliteit van de input (van de secundaire bronnen) wordt beïnvloed.

### Gestructureerde studie van datakwaliteit

Bij het bepalen van de kwaliteit van de gegevens in administratieve bronnen dienen veel kwaliteitsaspecten gemeten te worden. Tabel 5 geeft hiervoor in totaal 33 meetmethoden weer. Het lijkt dan ook niet erg efficiënt voor een NSI om de waarde van elk van die meetmethoden bij elke levering van een bron te bepalen. Zeker niet wanneer bronnen ‘stukje bij beetje’ geleverd worden, zoals bijvoorbeeld bij de levering van BTW-gegevens door de Belastingdienst aan het CBS het geval is. Het is dan ook aan te raden om bij het ontvangen van secundaire gegevens te beginnen met het bepalen van een beperkte set van essentiële, absoluut noodzakelijke, indicatoren. Wanneer daarbij geen problemen worden gevonden dient de data specifiek, meer gedetailleerd, bestudeerd te worden.

Een overzicht van de aanpak die uit de hierboven beschreven pragmatische aanpak volgt is in figuur 1 weergegeven. Het figuur bevat, naast een voorstel voor een stapsgewijze aanpak van het meten van de kwaliteit van de data, ook de eerder besproken checklist voor Bron en Metadata. Dit is gedaan om het overzicht compleet te maken. Daarnaast is ook het onderscheid in input- en outputkwaliteit in de figuur opgenomen. De drie stappen die bij het bepalen van de datakwaliteit onderscheiden worden zijn: i) Technische controles, ii) Nauwkeurigheid gerelateerde indicatoren, en iii) Output gerelateerde indicatoren. Elke stap wordt in de volgende paragrafen nader besproken.

**Figuur 1. Overzicht van het voorgestelde proces van kwaliteitscontrole voor secundaire bronnen op het CBS**



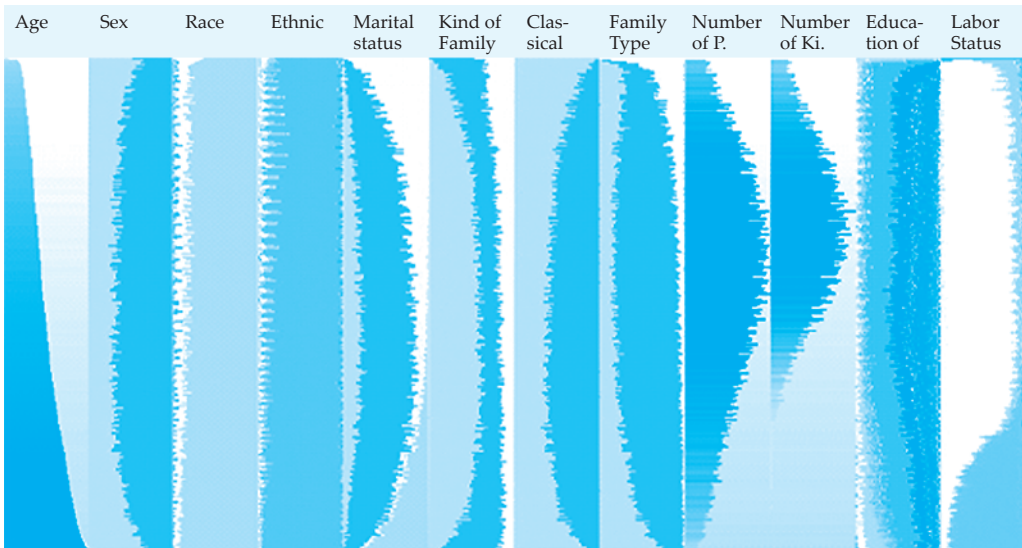
### Stap 1: Technische controles

In de eerste stap, die van de technische controles, worden snelle controles op (net) ontvangen gegevens uitgevoerd. Deze controles zijn op de input gericht, erg basaal, dienen snel te kunnen worden uitgevoerd en dienen uitsluitend te worden gebruikt om serieuze problemen te identificeren. Voorbeelden van dergelijke controles zijn: het vergelijken van de omvang van een bestand en/of het aantal (unieke) eenheden in een levering ten opzichte van die van eerdere leveringen en het technisch valideren van een leveringsbestand. In dit laatste geval wordt gekeken of de gegevens en opmaak van de gegevens in een levering aan de metadata-definitie voldoen. Dit is met name bij XML-bestanden een zeer gebruikelijke controle (Van der Vlist, 2002).

Een interessante toevoeging aan de controles in deze stap van het proces zijn controlemethoden waarin gebruik wordt gemaakt van grafische visualisaties. Doordat hedendaagse computers steeds sneller worden, meer grafische mogelijkheden bieden en steeds beter met (steeds) grote(re) databestanden om kunnen gaan wordt het gebruik van dergelijke methoden steeds aantrekkelijker. Figuur 2 laat een voorbeeld zien van het gebruik van een gevisualiseerde controle van een groot databestand. De figuur is afkomstig uit Theus (2006) en toont een 'table plot' van geaggregeerde gegevens van 12 variabelen. De gegevens zijn afkomstig uit een bestand met censusgegevens van de Verenigde Staten. Elke kolom geeft één variabele weer en elk 'meetpunt' is een samenvoeging van 250 waarnemingen. De gegevens zijn gesorteerd naar leeftijd; de variabele in de eerste kolom. Bij categoriale variabelen is voor elke antwoordcategorie een andere kleur gebruikt. Een 'table plot' wordt gebruikt om een totaal overzicht te geven van alle gegevens in een dataset. Wat in figuur 2 opvalt, is dat de verhouding mannen en vrouwen (in de 2e kolom) afwijkt voor de lage en hoge leeftijdsgroepen in het bestand. Voor de lage leeftijdsgroepen is dat opvallend te noemen omdat te verwachten is dat deze verhouding dicht bij die van de geboorte zal liggen (1,05:1; CIA, 2010).



Figuur 2. Een 'table plot' van geaggregeerde gegevens van de eerste 12 variabelen in Amerikaanse censusgegevens (uit Theus, 2006)



Een ander voorbeeld waarbij het gebruik van visualisatie methoden -aan de inputkant van het proces- nuttig kan zijn is het weergeven van ontbrekende gegevens. Vaak worden deze voor de gehele bron of per variabele getoond. Met behulp van visualisatie technieken kan de relatie tussen het ontbreken van gegevens voor meerdere variabelen bekeken worden. Een voorbeeld hiervan is te vinden in het artikel van Templ en Filmzner (2008). Daarnaast is het ook denkbaar de patronen van de ontbrekende gegevens te visualiseren. Dit soort patronen kan bijvoorbeeld anders zijn voor twee bronnen waarin dezelfde hoeveelheid gegevens ontbreken maar voor een ander aantal variabelen. Bijvoorbeeld wanneer in de eerste bron 10 000 gegevens van variabelen (min of meer) willekeurig ontbreken en in de tweede bron de gegevens van 100 variabelen voor exact 100 eenheden ontbreken. Het eerste geval is waarschijnlijk geen reden om contact op te nemen met de bronhouder en het tweede geval wel!

Mede omdat visualisatiemethoden ook bij datamining gebruikt worden (Pyle, 1999) is het zeer waarschijnlijk dat deze aanpak ook op de kwaliteitscontrole van secundaire databronnen toepasbaar is. Een mogelijk beperking zou kunnen zijn dat de visualisatie techniek (d.w.z. de weergave) die wordt gebruikt specifiek voor elke databron moet worden aangepast. Het gebruik van visualisatiemethoden voor de statistiek is iets dat serieus onderzocht moet worden. Indien dergelijke methoden inderdaad toepasbaar blijken dan is het belangrijk dat standaard methodieken ontwikkeld moeten worden om het gebruik hiervan te stimuleren. Groot voordeel van visualisatiemethoden is de mogelijkheid om alle gegevens in een bron snel te kunnen controleren. Dit is iets dat niet alleen voor de statistiek maar ook voor andere wetenschapsgebieden nuttig is.

### Stap 2: Nauwkeurigheid gerelateerde indicatoren

Wanneer voor een NSI duidelijk is voor welke publicatie een bron gebruikt gaat worden, kunnen meer specifieke kwaliteitscontroles worden toegepast. Indicatoren die hiervoor worden gebruikt zijn nog steeds input georiënteerd omdat ze naar de bruikbaarheid van de gegevens aan het begin van het proces kijken. Ze worden 'nauwkeurigheid gerelateerde' indicatoren genoemd omdat deze indicatoren, direct of indirect, aan de nauwkeurigheid van de gegevens gerelateerd zijn. Veel van de indicatoren die in tabel 5 staan vermeld behoren tot deze groep. Voorbeelden van 'nauwkeurigheid gerelateerde' indicatoren voor eenheden zijn: over- en onderdekking, selectiviteit en koppelbaarheid. Voorbeelden van dergelijke indicatoren voor variabelen zijn: selectiviteit, het percentage gaafgemaakte en geïmputeerde waarden en externe controle.

Veel van de indicatoren die in tabel 5 zijn weergegeven worden op het CBS al gebruikt voor het bepalen van de kwaliteit van enquêtegegevens. Dit vormt een goed startpunt voor het ontwikkelen van indicatoren die zowel voor primaire (enquête) als voor secundaire (administratieve) gegevens te gebruiken zijn. Dit uitgangspunt is een nobel streven maar op dit moment is nog onduidelijk in hoeverre dit in de praktijk mogelijk is.

### Stap 3: Output gerelateerde indicatoren

De kwaliteitsindicatoren die tot de derde stap van het bepalen van de datakwaliteit behoren, duiden de kwaliteit van administratiegegevens aan op geaggregeerd niveau. De indicatoren in deze groep zijn allen duidelijk output georiënteerd. Ze rapporteren op een niveau van kwaliteit dat met de inhoud van de geproduceerde statistieken samenhangt. Met de indicatoren in de 'aan output gerelateerde' groep wordt de vraag "hoe goed is de kwaliteit van de publicatie die is gebaseerd op deze set van gegevens?" beantwoord. Voorbeelden van indicatoren die tot de groep van 'aan output gerelateerde' indicatoren behoren zijn indicatoren die proberen de precisie van kernvariabelen te bepalen en indicatoren die proberen de selectiviteit van samengestelde totalen te bepalen. Onderzoek naar precisie-indicatoren heeft recentelijk aanzienlijk voortgang geboekt wanneer het schattingen betreft die op de combinatie van register- en enquêtegegevens zijn gebaseerd (Harmsen et al., 2009). De bepaling van de precisie van schattingen die volledig op administratieve gegevens zijn gebaseerd vormt echter nog steeds een grote uitdaging (Zhang, 2009). Ook de studie naar de selectiviteit van eenheden, het tweede voorbeeld, heeft de afgelopen jaren door het 'Representative Indicators for Survey Quality (RISQ)' project (Schouten et al., 2008) aanzienlijk voortgang geboekt. Hoewel de representativiteitindicatoren oorspronkelijk voor enquêtegegevens ontwikkeld zijn blijken ze ook toepasbaar op de eenheden in administratieve bronnen (Ouwenhand et al., 2009).

Er is echter ook een belangrijke algemene beperking in de toepasbaarheid van indicatoren die tot de derde stap behoren; mogelijk is dit ook bij -een deel van- de

indicatoren in de 2e stap het geval. Het is belangrijk dat de kwaliteitsindicatoren in de 3e stap algemeen toepasbaar zijn. Zeer specifieke indicatoren kunnen niet worden opgenomen omdat het eenvoudigweg niet mogelijk is alle denkbare indicatoren op te nemen (Daas et al., 2008). Dit belangrijk punt wordt door het volgende voorbeeld verduidelijkt. Een zeer specifieke indicator is bijvoorbeeld een indicator die de resultaten voor de schatting van 'het percentage werkloze personen in Nederland' (voor een bepaalde maand) afkomstig uit een administratieve bron en uit enquêtegegevens vergelijkt. Het moge duidelijk zijn dat een dergelijke indicator nuttig is maar niet zo gedetailleerd kan en mag worden opgenomen in de 3e stap van de Data hyperdimensie omdat deze indicator niet algemeen toepasbaar is. Om die reden is dan ook besloten de algemene indicator 'externe controle' (nr. 7.1) in tabel 5 te vermelden. Daarnaast is het ook nog eens zo dat verschillende gebruikers van een bron deze voor verschillende doeleinden willen gaan gebruiken waardoor ze elk weer andere eisen stellen aan de kwaliteit van de gegevens in de bron. Dit maakt het noodzakelijk de set van indicatoren die in de 3e stap van de hyperdimensie Data kan worden opgenomen enigszins te beperken.

Een andere belangrijke vraag voor de indicatoren in de 3e stap is de vraag of ze informatie over de bruikbaarheid van de bron leveren voor- of nadat de gegevens in de bron bewerkt zijn? Indien het indicatoren betreft die alleen na het bewerken van de gegevens kwaliteitinformatie leveren dan kun je je afvragen of deze indicatoren wel zo geschikt zijn om als indicatoren voor de kwaliteit van een bron te fungeren. Procesmatig gezien is een indicator die zo vroeg mogelijk in een proces een indicatie over de kwaliteit geeft te prefereren. Wat dat betreft lijken de controles en indicatoren in de eerste 2 stappen van het bepalen van de datakwaliteit veel beter geschikt. Hierbij is zeker sprake van het meten van kwaliteit aan het begin van het proces.

### *Het resultaat van het meten van datakwaliteit*

Het werk dat is uitgevoerd naar de kwaliteitsindicatoren in de hyperdimensie Bron en Metadata laat tevens zien dat het voor de toepasbaarheid van dat werk belangrijk is ook een gestructureerde manier van meten te ontwikkelen. Voor de Bron en Metadata hyperdimensie is daarvoor een checklist ontwikkeld (Daas et al., 2009b). Mede doordat de meetmethoden in beide hyperdimensies uit vrijwel uitsluitend kwalitatieve vragen bestaan was de keuze voor een checklist als hulpmiddel ook relatief eenvoudig.

Voor het gestructureerd meten van de kwaliteitsaspecten die tot Data hyperdimensie behoren, lijkt een checklist niet de meest handige oplossing. Zo zullen in de eerste, technische controle, stap visuele inspecties worden uitgevoerd terwijl daarnaast in de tweede en eventuele derde stap een aanzienlijk hoeveelheid kwantitatieve indicatoren moeten worden gemeten. De ontwikkeling van standaard scripts en/of een software programma lijken voor de Data hyperdimensie dan ook veel betere manieren om de gebruiker te ondersteunen.

Het is belangrijk dat de meetresultaten voor de indicatoren en controles in de hyperdimensie Data in een enkele rapportage worden samengebracht. Deze rapportage, die eerder Kwaliteitskaart of kwaliteitsinstrument is genoemd (Daas et al., 2008), dient algemeen toepasbaar te zijn en een duidelijk gestructureerd overzicht te geven van alle kwaliteitsaspecten die essentieel zijn voor de gegevens in de betreffende databron. Een andere uitdaging is de normering van de resultaten. Het vaststellen van normen voor de kwantitatieve resultaten van de indicatoren in Data zal een stuk lastiger zijn dan die voor de meer kwalitatieve kwaliteitsgegevens van de Bron en Metadata hyperdimensie (Daas et al., 2009a).

### 3. *Conclusie*

Dit document geeft een overzicht van de huidige stand van zaken en de ideeën voor de bepaling van de statistische bruikbaarheid van registers en administratieve bronnen op het CBS. Het eerste deel van het onderzoek naar het bepalen van de kwaliteit van dergelijke bronnen is recentelijk op het CBS afgerond. Dit werk heeft geresulteerd in de ontwikkeling van een raamwerk, dat uit drie hyperdimensies bestaat, en een checklist om de kwaliteitsindicatoren in de eerste twee hyperdimensies, Bron en Metadata, te bepalen. Indicatoren voor de bepaling van de kwaliteit van de gegevens van administratieve bronnen en registers treffen we in de Data hyperdimensie aan. Deze hyperdimensie wordt momenteel onderzocht. Dit document bevat een voorstel voor het op een gestructureerde manier bepalen van de kwaliteit van deze gegevens. Dit voorstel zal in het in april 2010 gestarte onderzoekproject BLUE Enterprise and Trade Statistics (BLUE-ETS), een project gefinancierd door het zevende kaderprogramma voor onderzoek van de Europese Unie, nader worden uitgewerkt. Hierbij zijn, naast de auteurs van dit document, ook statistici van de NSI's van Italië, Noorwegen, Slowakije en Zweden betrokken. Uiteindelijk doel van dit onderzoek is het opleveren van een instrument dat het mogelijk maakt de statistische bruikbaarheid van administratieve bronnen en registers op een efficiënt en eenduidige wijze te bepalen.

### *Referenties*

Bakker, B.F.M., Linder, F., and Van Roon, D. (2008). *Could that be true? Methodological issues when deriving educational attainment from different administrative datasources and surveys*. Proceedings of IAOS Conference on Reshaping Official Statistics, Shanghai, China.

Batini, C. and Scannapieco, M. (2006). *Data Quality: Concepts, Methodologies and Techniques*. Berlin: Springer.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., and Vignoles, A. (2010) *Widening Participation in Higher Education: Analysis Using Linked Administrative Data*. DoQSS Working Papers 1008, Department of Quantitative Social Science – Institute of Education, University of London.

CIA (2010) *The World Factbook, United States, People, Sex-ratio*, Online at <https://www.cia.gov/library/publications/the-world-factbook/geos/us.html>.

Daas, P.J.H., Arends-Tóth, J., Schouten, B., and Kuijvenhoven, L. (2008). *Quality Framework for the Evaluation of Administrative Data*. Proceedings of Q2008 European Conference on Quality in Official Statistics. Statistics Italy and Eurostat, Rome.

Daas, P.J.H., Ossen, S.J.L., and Arends-Tóth, J. (2009a). *Framework of Quality Assurance for Administrative Data Sources*. Paper for the 57th session of the International Statistical Institute, Durban, South Africa.

Daas, P.J.H., Ossen, S.J.L., Vis-Visschers, R.J.W.M., and Arends-Toth, J. (2009b). *Checklist for the Quality evaluation of Administrative Data Sources*. Discussion paper 09042, Statistics Netherlands, The Hague/Heerlen, The Netherlands.

Harmsen, C., Van Der Laan, J., and Kuijvenhoven, L. (2009). *Deriving longitudinal consistent household statistics from register information*. Paper for the 57th session of the International Statistical Institute, Durban, South Africa.

Karr, A. F., Sanil, A. P., and Banks, D. L. (2006). Data quality: A statistical perspective. *Statistical Methodology*, 3, 137–173.

Ouwehand, P., Schouten, B., and De Heij, V. (2009). *Representativity indicators for business surveys based on population totals*. Paper for the European Establishment Statistics Workshop, Stockholm, Sweden.

Pyle, D. (1999). *Data preparation for data mining*. San Francisco: Morgan Kaufmann.

Schouten, B., Cobben, F., and Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, 35, 101–113.

Schulte Nordholt, E. (2004). Introduction to the Dutch virtual census of 2001. In *The Dutch Virtual Census of 2001, analysis and methodology*, eds. E. Schulte Nordholt, M. Hartgers, R. Gircour, Voorburg: Statistics Netherlands.

Templ, M. and Filzmoser, P. (2008). *Visualization of missing values using the R-package VIM*. Forschungsbericht CS-2008-1, Institut f. Statistik u. Wahrscheinlichkeitstheorie, Wien, Austria.

Theus, M. (2006). Statistical Graphics. In *Graphics of Large Datasets: Visualizing a Million*, eds. A. Unwin, M. Theus, H. Hofmann, Singapore: Springer, pp. 31-54.

Unece (2007). *Register-based statistics in the Nordic countries – Review of best practices with focus on population and social statistics*. Geneva: United Nations Publication.

Van der Vlist, E. (2002). *XML Schema*. Sebastopol: O'Reilly & Associates.

Wallgren, A. and Wallgren, B. (2007). *Register-based Statistics: Administrative Data for Statistical Purposes*. Chichester: John Wiley & Sons.

Zhang, L-C. (2009). *Unit errors in statistical registers and their effects*. Paper for the 57th session of the International Statistical Institute, Durban, South Africa.

# *Het gebruik van BTW voor maand- en kwartaalomzetschattingen in Europa: Een vergelijking tussen Nederland en het Verenigd Koninkrijk*

Pieter Vlag, Henk van de Velden, Nino Mushkudiani (CBS)

## *1. Inleiding*

Het doel van dit onderzoek is om een vergelijking te maken in het gebruik van BTW-data voor maand- en kwartaalomzetschattingen tussen de verschillende Europese landen. Uiteindelijk doel hiervan is om de meest gebruikte methodes te beschrijven en aanbevelingen te doen over schattingen met (incomplete) registers voor de korte termijn statistieken (KTS). De aandacht gaat op dit moment uit naar de maand- en kwartaalomzetschattingen met BTW-data. Op termijn wordt ook een vergelijking gemaakt in het gebruik van registerdata over de sociale zekerheid voor de werkgelegenheidschattingen van de KTS. Dit paper beschrijft het eerste deel van het onderzoek: een vergelijking tussen Nederland en het Verenigd Koninkrijk

Voor de KTS is een groot verschil tussen data uit enquêtes en administratieve bronnen dat statistische bureaus (min of meer) controle hebben over de tijdigheid en opgaveperiode van enquêtedata. Voor administratieve data is dit niet het geval. Een gevolg hiervan is dat maand- en kwartaalschattingen noodgedwongen worden gemaakt op basis van incomplete BTW-registers. Ook verschilt de beschikbaarheid van de BTW-registers sterk per land. Door de grote onderlinge verschillen in de beschikbaarheid van BTW-data was de eerste stap van het onderzoek om na te gaan of het überhaupt wel mogelijk is om op Europees niveau aanbevelingen te doen voor het schatten van KTS-omzetontwikkelingen op basis van BTW.

We zijn begonnen met een case studie naar de bruikbaarheid van BTW-data voor maand- en kwartaalomzetschattingen in Nederland (NL) en het Verenigd Koninkrijk (VK). Het eerste doel van deze studie was om de grootste overeenkomsten en verschillen te identificeren. Om na te gaan of deze overeenkomsten en verschillen representatief zijn voor Europa is in het voorjaar van 2010 besloten om zes extra landen aan het onderzoek toe te voegen. Eveneens is toen een raamwerk ontwikkeld om de overeenkomsten en verschillen tussen de verschillende landen systematisch te onderzoeken. Dit raamwerk zal in 2010–2011 worden gebruikt door Nederland, Duitsland, Estland, Finland, Italië, Litouwen en het Verenigd Koninkrijk.

In deze bijdrage wordt de situatie tussen NL en VK vergeleken. Ook wordt het raamwerk gepresenteerd. De vergelijking met Duitsland, Estland, Finland, Italië, Litouwen wordt niet besproken omdat die nog in ontwikkeling is.

## 2. *Nederland en het Verenigd Koninkrijk*

Het Centraal Bureau voor de Statistiek (CBS) maakt op dit moment in een prototype al gebruik van BTW-omzetgegevens voor het maken van kwartaalomzetschattingen. De uitkomsten van dit systeem worden vergeleken met de kwartaalomzetschattingen uit het reguliere proces dat nog is gebaseerd op enquêtes. De benodigde BTW-gegevens worden geleverd door de Belastingdienst (BD).

Het huidige prototype wordt op dit moment omgezet naar een productiesysteem voor de KTS-omzetschattingen. Met dit nieuwe productiesysteem worden de huidige enquêtes vervangen door de kwartaalomzetschattingen op basis van de BTW-gegevens. Het idee was om BTW-maandopgaves te gaan gebruiken voor KTS-maandomzetschattingen. Het is echter momenteel de vraag of maandschattingen nog steeds gebaseerd kunnen worden op de BTW-maandaangevers, omdat door een wetsverandering het aantal maandaangevers sterk is afgenomen. Daarom zoekt het CBS voor maandstatistieken naar alternatieven.

Het statistisch bureau van het Verenigd Koninkrijk (Office of National Statistics, ONS) is nog bezig met de overweging om het register van de belastingdienst (Her Majesty Revenue and Customs, HMRC) te gebruiken voor haar KTS-schattingen. ONS heeft al wel enig voorbereidend onderzoek gedaan naar het gebruik van het BTW-register. Het is echter duidelijk dat Nederland verder is in het gebruik van BTW-gegevens voor KTS-omzetschattingen dan het VK.

## 3. *Het gebruik van administratieve bronnen voor de KTS in Nederland*

In deze paragraaf volgt een korte beschrijving van de huidige Nederlandse situatie en de toekomstige plannen voor het gebruik van BTW-gegevens voor KTS-omzetschattingen.

### 3.1 *Achtergrond huidige situatie*

Het verminderen van de administratieve lasten voor bedrijven is een belangrijk onderwerp in de Nederlandse politiek. De enquêtes waarmee de KTS-omzetschattingen gemaakt worden staan daarom continu onder politieke druk. Om de administratieve lasten die door het CBS worden veroorzaakt te verminderen heeft het CBS een methode ontwikkeld voor het maken van KTS-omzetschattingen die niet gebaseerd is op enquêtes maar op basis BTW-gegevens. De ingebruikname van



deze methode is echter tot twee keer toe negatief beïnvloed door wetswijzigingen over de BTW-aangifte.

Volgens de huidige plannen wordt een systeem voor het bepalen van de kwartaalomzetschattingen op basis van BTW-omzetdata met ingang van 2011 in reguliere productie genomen. Een voorloper van dit systeem wordt gebruikt om een belangrijke bron voor de eerste economische groeiraming, namelijk de reguliere kwartaalomzetschattingen verkregen met enquêtes, te valideren met BTW-omzetschattingen. De volgende eigenschappen van het Nederlandse BTW-register zijn belangrijk voor het gebruik voor statistische doeleinden:

1. De BTW mag op maand-, kwartaal- of op jaarbasis worden gerapporteerd.
2. Kleine bedrijven mogen BTW aangeven op jaarbasis, dat wil zeggen één keer per jaar voor het gehele kalenderjaar. De bijdrage van deze bedrijven aan de totale omzet is echter marginaal (<0.3% voor alle bedrijven in 2007). Omdat de bijdrage van de jaaraangevers op de totale omzet klein is, corrigeren we de maand- en kwartaalomzetschattingen hier niet voor.
3. De BTW-aangifte moet elektronisch gedaan worden binnen 30 dagen na het eind van de referentieperiode (maand, kwartaal of jaar).
4. In de praktijk blijkt dat de meeste bedrijven de BTW-aangifte tussen de 20ste en 30ste dag doen. Op dag 27 na het eind van de referentieperiode heeft grofweg 60–70% van de bedrijven aangifte gedaan.
5. Tussen de 20ste en 30ste dag worden BTW-registerdata dagelijks elektronisch van de Belastingdienst naar het CBS gestuurd. Er is dus slechts een kleine vertraging tussen BTW-aangifte en de beschikbaarheid over deze BTW-aangifte door het CBS.
6. Fiscale eenheden (bedrijven volgens de belastingdienst) kunnen verschillen van de statistische bedrijfseenheden van het CBS. Sinds de invoering van een nieuw bedrijvenregister beperkt deze ‘eenhedenproblematiek’ zich voornamelijk tot de grootste bedrijven.
7. De KTS-populatie wordt volledig door de BTW gedekt, gezien wettelijke verplichtingen en onderzoeksresultaten die bevestigen dat deze bedrijven daadwerkelijk aangifte doen.
8. Het CBS neemt aan dat voor de meeste bedrijfssectoren binnen het KTS-domein BTW-omzetontwikkelingen kunnen worden gebruikt voor reële omzetontwikkelingen.
9. Het CBS publiceert de KTS-resultaten doorgaans tussen de 45 en 60 dagen na het eind van de referentieperiode (maand of kwartaal). Alleen voor de sectoren “industrie” en “detailhandel” worden de schattingen eerder gepubliceerd, te weten na 30 dagen.

### 3.2 *Veranderingen in BTW-wetgeving*

In 2009 hebben er twee wijzigingen in de BTW-wetgeving plaatsgevonden (Tabel 1). Met ingang van januari 2009 werd de drempel voor verplichte aangifte op

maandbasis verhoogd van € 7 000,= naar € 15 000,=. Dit wil zeggen dat als de af te dragen BTW hoger was dan € 15 000,= per kwartaal de aangifte op maandbasis diende te worden gedaan. Door deze maatregel is het aantal maandaangevers afgenomen. De daaropvolgende wijziging was de afschaffing van de verplichte maandaangifte m.i.v. het derde kwartaal van 2009. Bedrijven mogen dus vanaf het 3<sup>e</sup> kwartaal 2010 zelf kiezen of ze per maand of per kwartaal aangifte doen.

**Tabel 1**  
**BTW-omzet en -afdracht 2008–2010**

	BTW: maandaangevers			BTW-kwartaalaangevers			BTW: maand- + kwartaalaangifte			dekking maandaangifte		
	aantal x 1 000	omzet fictieve eenheid	afdracht fictieve eenheid	aantal x 1 000	omzet fictieve eenheid	afdracht fictieve eenheid	aantal x 1 000	omzet fictieve eenheid	afdracht fictieve eenheid	aantal %	omzet %	afdracht %
<b>2008</b>												
kw III	291	463 152	14 849	856	76 442	505	1 147	539 594	15 354	25,4	85,8	96,7
kw IV	291	483 179	15 477	851	75 537	798	1 142	558 715	16 275	25,5	86,5	95,1
<b>2009</b>												
kw I	255	398 551	12 130	940	68 386	917	1 194	466 937	13 047	21,3	85,4	93,0
kw II	254	393 927	13 235	956	74 157	1 061	1 210	468 084	14 296	21,0	84,2	92,6
kw III	195	245 474	2 741	1 028	218 126	10 443	1 223	463 600	13 183	16,0	52,9	20,8
kw IV	193	263 857	2 665	1 030	256 022	13 203	1 223	519 879	15 868	15,8	50,8	16,8
<b>2010</b>												
kw I	196	253 586	1 498	1 054	232 506	11 318	1 250	486 092	12 816	15,7	52,2	11,7

Door deze wijzigingen is het aandeel van de omzet van de maandaangevers op de totale omzet afgenomen. In het derde kwartaal van 2008 dekten BTW-maandaangiften ruim 85% van de omzet. In het eerste kwartaal van 2010 was het aandeel van de omzet van maandaangevers echter nog slechts 50% van het totaal. We verwachten dat deze afname de komende jaren verder gaat. Een consequentie van deze afname is dat de representativiteit van de maandaangevers voor de totale populatie in het geding komt. Een aanwijzing hiervoor is de grote afname van de dekking maandaangevers in de BTW-afdracht dan in de BTW-omzet.

Tijdreeksen in omzetontwikkeling vanaf 3<sup>e</sup> kwartaal 2009 (= laatste wetwijziging) suggereren eveneens dat omzetten van BTW-maandaangevers niet meer representatief voor de populatie zijn. Nadeel van deze tijdsreeksen is dat zij nog te kort zijn om definitieve conclusies trekken. Echter de aanwijzigingen zijn zo sterk dat nu al de conclusie kan worden getrokken dat de aanname dat maandaangevers representatief zijn voor de gehele populatie niet zomaar mag worden gemaakt.

### 3.3 *Het gebruik van BTW-omzet voor kwartaalomzetschattingen*

Het prototype van het productiesysteem heeft tot veel inzichten in het gebruik van het BTW-register voor kwartaalomzetschattingen bij het CBS opgeleverd (o.a. op het punt van revisies t.o.v. definitieve uitkomsten en kwaliteit bedrijfstypering). Zoals eerder gemeld wordt dit systeem nu omgebouwd tot een productiesysteem. Dit nieuwe systeem moet in 2011 in productie worden genomen. Vanaf dat moment zal de huidige enquête niet meer worden gehouden.

In de nieuwe situatie zal het CBS de grootste 1900 bedrijven in Nederland, (de zogenaamde TOP-1900 bedrijven) primair blijven waarnemen. De BTW-omzet zal gebruikt worden voor de overige bedrijven. Voor sommige bedrijfsklassen zullen de kwartaalomzetschattingen in zijn geheel gebaseerd zijn op BTW-gegevens omdat niet in elke bedrijfsklasse een TOP 1900 bedrijf aanwezig is.

De methode voor het schatten van kwartaalomzetten (45–60 dagen na het eind van de referentieperiode) met behulp van BTW-data is relatief eenvoudig, aangezien de BTW-aangiften op dat moment bijna in zijn geheel binnen zijn. De omzetgroei is op dat moment simpel te berekenen door de omzet van alle bedrijven in periode  $t$  te delen door de omzet van alle bedrijven in periode  $t-4$  (hetzelfde kwartaal een jaar eerder).

Voor de snelle kwartaalschattingen (gezien de verwerkingstijd binnen het CBS moeten deze zijn gebaseerd op BTW-data die 27 dagen na afloop van de referentieperiode binnen zijn) moest een methode ontwikkeld worden voor de nog ontbrekende BTW-data. Besloten is om deze ontbrekende data te imputeren. Voor het imputeren is het van belang om de opgaveperiode van de BTW-aangiften te weten. Met andere woorden: doet de eenheid op maand- of op kwartaalbasis aangifte en is er een verandering in aangifteperiode ten opzichte van vorig jaar? Ook hier negeren we voor het gemak de jaaraangevers. Een eenheid valt daarom altijd in een van de volgende acht aangiftegroepen: MM, MQ, QM, QQ, xM, xQ, Mx, Qx. Waarbij M voor maandaangever staat, Q voor kwartaalaangever, x voor een ontbrekende aangifte (bijv. in geval van een gestart en gestopt bedrijf). Verder staat in deze systematiek de tweede letter voor de periode  $t$  (= huidig kwartaal) en de eerste letter voor periode  $t-12$  (= hetzelfde kwartaal in voorgaand jaar).

Na het vaststellen van deze acht groepen worden voor bedrijven met dezelfde bedrijfsactiviteit de ontbrekende waarden geïmputeerd met de gemiddelde groeivoet (=omzet  $t$ /omzet  $t-12$ ) per groep van de beschikbare BTW-data (zie Vlag en van den Bergen, 2010). Impliciet is dus de aanname dat binnen een groep de snelle aangevers (= BTW-data binnen 27 dagen na afloop kwartaal) qua groeivoet representatief zijn voor de gehele populatie. Onder deze aanname worden immers groeivoeten voor de mindere snelle aangevers (= BTW-data nog niet binnen 27 dagen na afloop kwartaal) geïmputeerd.

Voor de groepen Mx en Qx (=stoppende bedrijven) wordt de waarde 0 geïmputeerd. Ontbrekende waarden in de groepen xM en xQ (=startende bedrijven) worden geïmputeerd met het gemiddelde omzetniveau van de beschikbare BTW-data.

Impliciet is dus de aanname dat voor starters de snelle aangevers (= BTW-data binnen 27 dagen na afloop kwartaal) qua omzetniveau representatief zijn voor de gehele populatie (per bedrijfsactiviteit). Onder deze aanname worden immers de omzetten van de minder snelle aangevers (als de BTW-data nog niet binnen zijn na 27 dagen na afloop van het kwartaal) geïmputeerd. Deze methode wordt met zijn voor- en nadelen behandeld in Vlag en van den Bergen, 2010. In dit rapport worden eveneens de resultaten besproken.

Wordt voor snelle kwartaalschattingen de impliciete aanname gedaan dat de beschikbare BTW-data representatief zijn voor de gehele populatie, voor de maandschattingen kan deze aanname niet worden gebruikt. Voor maandschattingen zijn immers alleen gegevens van BTW-maandaangevers beschikbaar en in paragraaf 3.2 is juist beargumenteerd dat de aanname dat BTW-maandaangevers representatief zijn voor de gehele populatie sinds de recente wetswijzigingen niet zomaar meer mag worden gemaakt.

Afhankelijk van de toekomstige ontwikkelingen blijven er dan drie mogelijkheden voor het maken van maandschattingen:

1. Een enquête onder de grootste bedrijven en een modelmatige schatting gebaseerd op het (gemiddelde) historische verschil tussen de groeivoeten van de grootste bedrijven en de totale populatie (met een bepaalde bedrijfsactiviteit). Hierbij wordt geen gebruik meer gemaakt van BTW-maandgegevens. In Finland worden eveneens dergelijke methodes ontwikkeld.
2. Een enquête onder de grootste bedrijven in combinatie met een aangepaste versie van de methode “Vlag en van den Bergen”, waarin gecorrigeerd wordt voor niet-representatieve BTW-maandgegevens.
3. Een enquête bij de grootste bedrijven gecombineerd met een kleine enquête bij kleinere bedrijven. Schattingen op een lager aggregatieniveau worden gemaakt aan het eind van het kwartaal als de BTW-aangiften beschikbaar zijn. Deze methodiek wordt in Zweden toegepast (Lennartsson en Lindblom, 2009)

## **4. *Het gebruik van administratieve bronnen voor de KTS in het Verenigd Koninkrijk***

### **4.1 *Achtergrond huidige situatie VK***

Het Office for National Statistics (ONS) in het Verenigd Koninkrijk heeft nog geen concrete plannen om BTW-gegevens te gaan gebruiken voor de KTS-omzet schattingen. Er wordt echter nu gekeken naar de mogelijkheden om BTW-registerdata te gaan gebruiken voor de KTS (Orchard en James, 2010).

Op dit moment maakt ONS gebruik van drie enquêtes voor omzetcijfers. De drie enquêtes dekken verschillende bedrijfssectoren, te weten detailhandel, industrie, distributie (transport) en zakelijke diensten.

#### 4.2 *Het BTW-register*

ONS heeft (nog) geen directe toegang tot het BTW-register. Het register is eigendom van de belastingdienst (HMRC). De vier belangrijkste variabelen binnen dit register zijn: BTW-omzet, BTW-identificatienummer (het unieke identificatienummer van het bedrijf), rapportagefrequentie en rapportageperiode.

Net als in Nederland wordt in het VK de BTW op maand-, kwartaal- en jaarbasis gerapporteerd. Echter in het VK zijn er drie verschillende soorten kwartaalaangevers. De kwartaalaangevers mogen zelf kiezen of ze over de maanden 1–3, 4–6, 7–9, 10–12 (type 1 = kalenderkwartaal), de maanden 2–4, 5–7, 8–10, 11–1 (type 2 = één maand verschuiving t.o.v. het kalenderkwartaal) of de maanden 3–5, 6–8, 9–11, 12–2 (type 3 = twee maanden verschuivingen t.o.v. kalenderkwartaal) aangifte doen. Kwartaalaangiften in het VK zijn dus eigenlijk BTW-aangiften voor 3-maandsperiodes (Orchard, 2010). Deze opdeling in drie soorten kwartaalaangevers (type 1, type 2 en type 3) creëert een extra probleem voor het gebruik van BTW voor de KTS. De vraag die relevant is: hoe kunnen de drie typen kwartaal BTW-omzetdata worden opgesplitst in maandomzetten om ze na deze ‘opsplitsing’ weer te gebruiken voor maand- en kwartaalomzetschattingen?

In vergelijking met Nederland zijn er weinig BTW-maandaangevers in het VK. Het omzetaandeel van maandaangevers t.o.v. de totale jaaromzet is 15%. Echter de maandaangevers behoren tot een select aantal bedrijfsklassen en kunnen daarom niet als representatief worden beschouwd.

Qua aantallen zijn kwartaalaangevers type 1, type 2 en type 3 ongeveer even groot. Het omzetaandeel van kwartaalaangevers type 1 is echter aanzienlijk groter dan voor type 2 en 3. Hieruit kan geconcludeerd worden dat grote bedrijven hun aangifte volgens het financiële kwartaal doen. Er wordt nu onderzoek gedaan naar de representativiteit van type 1 kwartaalaangevers voor de twee overige type kwartaalaangevers.

Het omzetaandeel van maandaangevers is met 15% relatief hoog. Echter de maandaangevers behoren tot een select aantal bedrijfsklassen en kunnen daarom niet als representatief worden beschouwd.

#### 4.3 *Tijdigheid BTW-aangiften en publicatiemomenten*

In het VK moeten bedrijven binnen 40 dagen na de rapportageperiode aangifte doen. De belastingdienst in het VK stuurt de data slechts één keer per maand naar ONS. Hierdoor is de tijdigheid van de BTW-data in het VK een groot probleem: pas 118 dagen na het einde van de rapportageperiode beschikt ONS over 100% van de BTW-aangiften.

Echter met de huidige enquête methode worden de KTS erg snel gepubliceerd: de Productie Index (IoP) binnen 26 dagen, de Diensten Index (IoS) binnen 40 dagen en de Detailhandelsindex (RSI) tussen de 15 en 20 dagen. Dit betekent dat in de praktijk – als het VK daadwerkelijk besluit om administratieve data te gaan gebruiken voor de KTS – mogelijkerwijs zal moeten worden ingezet op modelmatige schattingen gebaseerd op:

- 1) het omzetten van de drie types kwartaalaangiften naar maandomzetten;
- 2) vervolgens gebruik te maken van het historische verschil in omzetcotwikkeling tussen de grootste bedrijven die nog steeds geënquêteerd worden en de totale populatie met dezelfde bedrijfsactiviteit. Voor bepaling van dit verschil kan gebruik worden gemaakt van de drie types kwartaalaangiften.

## 5. *Vergelijking NL en VK*

De grote verschillen in tijdigheid tussen NL en VK zijn:

1. Het CBS krijgt de BTW-data eerder dan ONS.
2. ONS produceert haar cijfers sneller.
3. In Nederland komt het kwartaal van een kwartaalaangever in 99% overeen met een kalenderkwartaal. In het VK rapporteren bedrijven niet per kalenderkwartaal maar per verschillende driemaandsperiodes.

Door de verschillende type driemaandsperiodes voor het aangeven in het VK is het onmogelijk om de Nederlandse methodiek over te nemen in het VK. In de volgende paragraaf laten we echter zien dat gedeeltes van het Nederlandse systeem mogelijk ook in het VK bruikbaar zijn nadat de kwartaalomzetten opgedeeld zijn in drie maandomzetten.

Het gegeven dat de maandaangevers in het VK niet representatief zijn voor de totale populatie is interessant voor Nederland. Als meer bedrijven in Nederland overstappen van een maand- naar een kwartaalaangifte dan worden de huidige aanwijzingen bevestigd dat Nederlandse maandaangevers ook niet meer representatief zijn voor de totale populatie.

## 6. *Toekomstig onderzoek in het VK*

Om de BTW-data geschikt te maken voor de KTS is ONS van plan om de volgende stappen te nemen:

1. De drie types kwartalen op te delen in maanden, gebaseerd op historische data en de enquête resultaten van de grootste bedrijven.
2. Afhankelijk van de representativiteit van de verschillende type kwartaalaangevers voor de totale populatie:
  - 2.1 Als de types niet representatief zijn: schatten op basis van historische gegevens van de BTW-omzet voor de huidige en vorige maand met tijdreeks analyses voor de BTW-typen welke de laatste maand niet hebben. Dit is het meest waarschijnlijke scenario.
  - 2.2 Als de types representatief zijn: een schattingsmethodiek zoals beschreven door Vlag en van den Bergen (2010) om bij te schatten voor berichtgevers die opgave doen volgens het type kwartaalaangiften dat niet de laatste maand dekt.

Als men de punten 2.1. en 2.2 op hoog abstractieniveau bekijkt dan lijkt het erop dat ondanks de grote verschillen in de beschikbaarheid van de data voor de KTS Nederland en het VK tegen dezelfde soort problematiek aanlopen.

Samenvattend, als het VK erin slaagt om de omzetten van verschillende types BTW- kwartaalaangevers om te rekenen naar maandomzetten, is het probleem voor het VK en Nederland hetzelfde. Als de ‘omgerekende’ BTW-omzetten representatief zijn voor de populatie kan een al dan niet aangepaste versie van de methode Vlag en Van den Bergen (2010) gebruikt worden. Als de ‘omgerekende’ BTW-omzetten niet representatief zijn voor de populatie willen beide landen historische data gebruiken om aan de hand van omzetontwikkelingen voor de grootste de omzetontwikkeling voor de gehele populatie te beschrijven.

## 7. *Conclusies en discussie*

De beschikbaarheid van BTW-data voor KTS-schattingen verschilt aanzienlijk tussen het VK en Nederland. De belangrijkste verschillen zijn:

1. Tijdigheid: 30 dagen na het eind van de rapportage periode in Nederland en 118 dagen na het eind van de rapportage periode in VK zijn de data binnen. Hierna volgen nog enkele nazendingen en correcties op opgaves.
2. Rapportageperiodes: maand, kwartaal en jaar in Nederland, maand-, drie verschillende type driemaands- en jaaraangevers in VK.

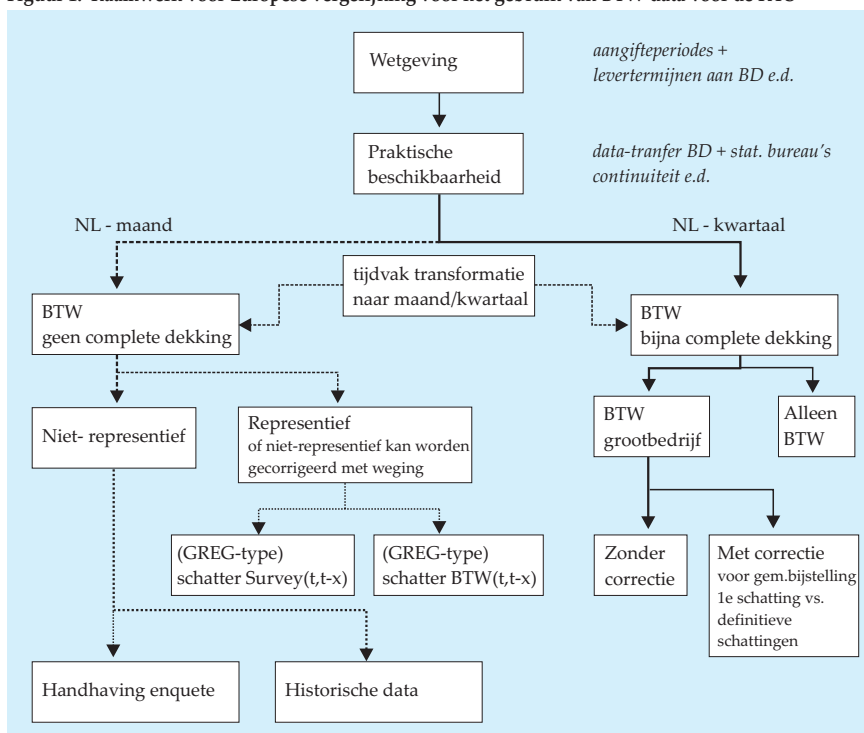
We kunnen daarom concluderen dat het Nederlandse systeem voor kwartaalaangevers niet gebruikt kan worden in het VK. Echter na het opdelen van de kwartalen in maanden van de verschillende type kwartaalaangevers is de problematiek voor maandaangevers in het VK en Nederland hetzelfde. De keuze die gemaakt dient te worden is dan:

1. Als de data representatief zijn voor de gehele populatie kan een systeem voor de KTS-schattingen worden opgezet onder de aanname dat de groeifactor van de ontbrekende BTW-data gelijk is aan de groeifactor van de beschikbare BTW-data (Vlag en van den Bergen, 2010).
2. Als de data niet representatief zijn voor de gehele populatie is de meest logische optie dat gebruikt wordt gemaakt van een model dat is gebaseerd op het historische verschil tussen ontwikkelingen tussen grote bedrijven en het midden- en kleinbedrijf.

Om tot aanbevelingen te komen over ‘best practices’ voor KTS-schattingen op incomplete administratieve data is mogelijk als er een relatie kan worden gelegd tussen beschikbare data en de praktijk.

Om de beschikbaarheid van register data en de gebruikte methodes in de verschillende landen systematisch met elkaar te kunnen vergelijken is door de samenwerkende statistiek bureaus een raamwerk ontwikkeld (zie Figuur 1).

Figuur 1. Raamwerk voor Europese vergelijking voor het gebruik van BTW-data voor de KTS



De gedachte achter dit raamwerk is dat afhankelijk van de wetgeving en data-transfer tussen de administratiehouders en statistische bureaus, het volgende onderscheid kan worden gemaakt bij het gebruik (en bijbehorende methodes) van administratieve data voor KTS-schattingen:

1. De administratieve data (bijv. BTW) dekken maar een beperkt deel van de populatie bedrijven op het moment dat de KTS-schattingen moeten worden gemaakt (linker blok in Figuur 1).
2. De administratieve data (bijv. BTW) dekken het grootste deel van de populatie bedrijven op het moment dat de KTS-schattingen moeten worden gemaakt (rechter blok in Figuur 1).

In het eerste geval (beperkte dekking) zijn er op dit moment drie benaderingen in Europa in gebruik of in ontwikkeling. Deze drie benadering zijn:

1. Eenn omzetschatting op basis van BTW-gegevens is niet mogelijk, enquête blijft noodzakelijk.
2. Schatten op basis van historische gegevens.
3. Gebruik maken van schattingsmodellen, bijvoorbeeld door directe schattingen of modellen van het regressietype zoals GREG. Deze laatste mogelijkheid kan opgedeeld worden in:



- 3.1 Het wegen van de enquêteuitkomsten voor de huidige periode t met behulp van complete administratieve data voor t-x.
- 3.2 Het wegen van incomplete administratieve data op tijdstip t gebruikmakend van de complete administratieve data van tijdstip t-x

De vierde benadering in Europa is in het geval van een bijna complete dekking:

4. KTS-schattingen worden alleen gebaseerd op administratieve data (BTW of werkgelegenheid) waarbij wordt aangenomen dat de groeivoet voor de beschikbare data gelijk is aan de groeivoet voor de (nog) niet beschikbare data.

Deze laatste mogelijkheid kan ook worden opgedeeld in:

- 4.1 Alleen bovenstaande aanpak
- 4.2 Bovenstaande aanpak met een revisiecorrectie, waarin de revisiecorrectie eruit bestaat door op macroniveau te gecorrigeren voor het gemiddelde verschil tussen de eersteschatting en de definitieve schatting (over de laatstgenoemde x maanden)

Doel van deze opzet is om uit te zoeken of zij eveneens hetzelfde 'pad' uit Figuur 1 kunnen volgen als de verschillende statistische bureaus een vergelijkbare dekking van de administratieve data hebben op het moment dat de schattingen moeten worden gemaakt. Mogelijk kunnen delen van de methodologische aanpak door deze bureaus kunnen worden gedeeld om tot een betere methode te komen. Dit raamwerk zal gebruikt worden om aanbevelingen te doen voor succesvolle methoden voor het gebruik van registers voor het maken van KTS-schattingen. Het raamwerk zal in 2010–2011 verder worden uitgewerkt door: Duitsland, Estland, Finland, Italië, Litouwen, het VK en Nederland door het beschrijven van de beschikbare registerdata en methoden voor KTS-schattingen op verschillende tijdstippen en het identificeren en oplossen van gedeelde problemen.

## Referenties

Y. Beuken en P. Vlag, 2010, *Business Register: the Dutch experience*, Statistics Netherlands, ESSnet seminar on admin data – Rome March 2010

D. Lennartsson and A.Lindblom, 2009, *The implementation of a more efficient way of collecting data – Turnover in other services*, final report MEETS-program: Grant agreement No 30102.2008.002–2008.369

C. Orchard, 2010, *File Preparation, Data Matching, and Distribution Analysis of VAT Turnover*, internal report ESSnet WP4.

C. Orchard and G. James, 2010, *The Use of VAT Turnover Data in Short Term Surveys Within the UK*, deliverable ESSnet AdminData.

H. van de Velden, P. Vlag and N. Mushkudiani, 2010, *Use of VAT in the Netherlands: a brief overview*, deliverable ESSnet AdminData

P. Vlag and D. van den Bergen, 2010, *The use of VAT for short term statistics: some quality aspects*, Statistics Netherlands, ESSnet seminar on admin data – Rome March 2010

# *Modellen voor micro-integratie*

Jeroen Pannekoek (CBS)

## **1. Inleiding**

Statistische bureaus maken meer en meer gebruik van gegevens afkomstig uit registraties. De grote voordelen van het gebruik van registraties ten opzichte van het gebruik van enquêtes zijn al vaak verwoord, de belangrijkste zijn: de beschikbaarheid van grote hoeveelheden data zonder daarvoor duur en tijdrovend veldwerk hoeven te verrichten en het niet (nogmaals) een beroep hoeven te doen op berichtgevers waardoor een reductie in de enquêtedruk ontstaat.

In veel gevallen zijn administratieve bronnen alleen echter niet voldoende om de gedetailleerde statistische publicaties samen te stellen die statistische bureaus moeten leveren en zijn gegevens uit steekproefonderzoek nodig om de ontbrekende informatie aan te vullen. In dit rapport beschouwen we bijvoorbeeld de situatie dat er voor het samenstellen van bedrijfsstatistieken verschillende bronnen beschikbaar zijn. Enkele belangrijke variabelen, zoals de omzet en het aantal werkzame personen zijn beschikbaar uit betrouwbare administratieve bronnen. Op basis van deze informatie kunnen, per maand en/of kwartaal, totalen voor deze variabelen gepubliceerd worden voor vele verschillende bedrijfstakken, regio's en grootteklassen. Voor meer gedetailleerde statistieken, zoals de jaarlijkse productiestatistieken, is het nodig om informatie te vergaren door steekproefonderzoek.

De gegevens uit het steekproefonderzoek kunnen veelal gekoppeld worden aan de administratieve bronnen. Er ontstaan dan records waarin variabelen uit verschillende bronnen voorkomen. Dit kan leiden tot inconsistenties als dezelfde variabele in verschillende bronnen niet dezelfde waarde heeft. Het kan ook leiden tot inconsistenties tussen verschillende variabelen afkomstig uit verschillende bronnen omdat de variabelen, vooral bij bedrijfsstatistieken, onderling verbonden zijn door allerlei logische en boekhoudkundige relaties. Om de inconsistenties tussen de variabelen in gekoppelde records op te lossen is een micro-integratie stap nodig.

In dit rapport beschouwen we formele modellen die geautomatiseerd toegepast kunnen worden om een micro-integratie stap uit te voeren waarbij de mogelijk conflicterende informatie uit de verschillende data bronnen en de logische en boekhoudkundige regels met elkaar in overeenstemming worden gebracht door aanpassingen aan de data. Dit soort methoden zijn traditioneel vooral toegepast op macro-niveau, d.w.z. op geaggregeerde data. Hier laten we zien hoe deze aanpak toegepast kan worden op verschillende integratieproblemen op micro-niveau.

In paragraaf 2 wordt de relatie geschetst tussen formele methoden voor het aanpas-

sen van data op macro-niveau en het aanpassen van data op micro-niveau. In paragraaf 3 worden een drietal aanpasmodellen besproken en in paragraaf 4 worden de verschillende opties die er zijn voor het toepassen van deze modellen toegelicht aan de hand van een voorbeeld. In paragraaf 5 worden tenslotte enkele conclusies samengevat.

## 2. *Aanpassen van data op macro- en micro-niveau*

### 2.1 *Klassieke toepassingen op macro-niveau*

Het aanpassen van data zodanig dat die data aan bepaalde gespecificeerde formele regels voldoen is traditioneel voornamelijk toegepast op geaggregeerde data, zoals tabellen met aantallen of totalen. Methoden op dit terrein zijn sinds ongeveer 1940 toegepast in de sociale wetenschappen en de macro-economie. In deze paragraaf geven we twee eenvoudige voorbeelden van aanpasproblemen op macro-niveau en zullen we vervolgens de overeenkomst laten zien met het aanpasprobleem op micro-niveau dat het onderwerp is van dit rapport.

**Tabel 1**  
Officiële schattingen (in duizendtallen) van de burgerlijke staat van de vrouwelijke bevolking van Engeland en Wales naar leeftijd in 1957

	Alleenstaand	Gehuwd	Gehuwd geweest	Totaal
<i>Leeftijd</i>				
15–19 jaar	1 306	83	0	1 389
20–24 jaar	619	765	3	1 387
25–34 jaar	436	2 566	37	3 039
35–44 jaar	330	2 765	132	3 227
45–49 jaar	208	1 350	108	1 666
50 jaar en ouder	1 116	4 100	2 329	7 545
Totaal	4 015	11 629	2 609	18 253

Een historisch voorbeeld, uit de sociale wetenschappen, van het aanpassen van een tabel is beschreven in Friedlander (1961) en heeft betrekking op de data getoond in Tabel 1<sup>1</sup>. Het betreft bevolkingsaantallen naar leeftijd en burgerlijke staat van vrouwen uit Engeland en Wales in 1957. Voor deze tabel kwamen in 1958 nieuwe schattingen voor de marginalen beschikbaar en Friedlander paste het binnenwerk van de 1957-tabel aan, zodanig dat de aangepaste tabel in overeenstemming was met de nieuwe marginalen en het binnenwerk van de aangepaste tabel en de oorspronkelijke tabel zo weinig mogelijk van elkaar verschilden.

Als we de rij- en kolomtotalen van de aangepaste tabel noteren met  $\tilde{X}_i$  en  $\tilde{X}_{.j}$ , en de nieuwe marginalen met  $\tilde{X}_i^N$  en  $X_i^N$ , dan kan het aanpasprobleem geformuleerd worden als

Minimaliseer  $D(\mathbf{x}, \tilde{\mathbf{x}})$

Onder de voorwaarden  $\tilde{X}_{.i} = X_{.i}^N$ ,  $\tilde{X}_{.j} = X_{.j}^N$  voor  $i = 1, \dots, 6; j = 1, \dots, 3$ ,

waarbij  $\mathbf{x}$  staat voor het binnenwerk van de oorspronkelijke tabel en  $\tilde{\mathbf{x}}$  voor het binnenwerk van de aangepaste tabel. De functie  $D$  geeft het verschil weer tussen  $\mathbf{x}$  en  $\tilde{\mathbf{x}}$ . De oplossing van deze minimalisatie,  $\tilde{\mathbf{x}}$ , hangt af van de keuze van de functie  $D$ .

De oplossing van Friedlander correspondeert met de keuze

$$D(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{ij} \tilde{x}_{ij} (\ln \tilde{x}_{ij} - \ln x_{ij}).$$

Deze functie  $D$  staat bekend als de "Kullback-Leibler divergence".

Dit criterium leidt tot een heel eenvoudig algoritme om de tabel aan te passen. Dat algoritme gaat als volgt, stap 1: vermenigvuldig iedere rij van het oorspronkelijke binnenwerk met een factor zodanig dat de resulterende waarden optellen tot de gewenste nieuwe rijtotalen. Stap 2: vermenigvuldig vervolgens iedere kolom van de nu gevormde aangepaste tabel met een factor zodanig dat de resulterende waarden optellen tot de gewenste nieuwe kolomtotalen. Herhaal deze stappen tot convergentie is bereikt. Dit eenvoudige algoritme, dat bekend staat als "iterative proportional fitting" (IPF) of ook wel RAS, is voor het eerst beschreven door Deming en Stephan (1940) en wordt nog steeds veelvuldig toegepast in verschillende situaties waarbij het aanpassen van tabellen aan nieuwe informatie aan de orde is. De in dit rapport toegepaste algoritmen kunnen in zekere zin als generalisaties van dit iteratieve algoritme worden opgevat, waarbij meer algemene lineaire gelijkheid- en ongelijkheidsrestricties meegenomen worden en andere doelfuncties  $D$  worden beschouwd (zie De Waal et al., 2011, hfdst. 10).

Bekende voorbeelden van aanpasproblemen uit de macro-economie zijn het aanpassen van input-output tabellen aan nieuwe marginalen en het consistent maken van aanbod- en gebruiktabellen (Stone et al., 1942). Een sterk vereenvoudigd voorbeeld van aanbod- en gebruiktabellen, met gefingeerde data, is weergegeven in tabel 2<sup>2</sup> (overgenomen uit Bikker et al., 2007). Er zijn hier slechts twee bedrijfstakken: industrie en dienstverlening en twee productgroepen: industrieproducten en diensten. Een cel in de aanbodtabel bevat de waarde in geld van de aangeboden producten in één van de productgroepen door één van de producenten. De eerste twee rijen in de gebruiktabel hebben betrekking op de afname van de verschillende producten door verschillende afnemers. De bedragen in de laatste twee rijen van deze tabel, Lonen en Exploitatieoverschot hebben betrekking op de overige bestedingen van de afnemers.

**Tabel 2 Aanbod- en gebruiktabellen**  
**Tabel 2a Aanbod**

	Producenten		
	Industrie	Dienstverlening	Totaal
	<i>x mln.</i>		
<i>Producten</i>			
Industrieproducten	700	300	1 000
Diensten	100	400	500
Totaal	800	700	1 500

**Tabel 2b**  
**Gebruik**

	Afnemers			
	Industrie	Dienstverlening	Consumenten	Totaal
	<i>x mln.</i>			
<i>Producten</i>				
Industrieproducten	50	190	860	1 100
Diensten	170	100	180	450
Lonen	450	350		800
Exploitatieoverschot	130	60		190
Totaal	800	700	1 040	2 540

De gegevens in deze twee tabellen moeten voldoen aan de volgende restricties:

1. Voor Industrie en Dienstverlening is het totale aanbod gelijk aan het totale gebruik (de betreffende kolomtotalen van tabel 2a zijn gelijk aan die van tabel 2b)
2. Voor Industrieproducten en Diensten is het totale aanbod gelijk aan het totale gebruik (de betreffende rijtotalen in tabel 2a zijn gelijk aan de overeenkomstige rijtotalen in tabel 2b)

Aan deze laatste restrictie is in dit voorbeeld niet voldaan. In de praktijk komt het vaak voor dat niet aan dit soort restricties wordt voldaan omdat de cijfers in cellen van deze tabellen schattingen zijn op basis van verschillende surveys en registraties. Door combinatie van de informatie uit de data en de restricties kunnen aangepaste aanbod- en gebruiktabellen geconstrueerd worden die wel aan de restricties voldoen en zo min mogelijk van de oorspronkelijke tabellen afwijken. Voor dit soort toepassingen is, bijvoorbeeld door Stone et al. (1942), voorgesteld om het minimaliseren van de gewogen kwadratische verschillen als criterium te nemen. Als we de totalen van de eerste twee rijen van tabel 2a en 2b weergeven met  $\tilde{X}_{a,i}$  en  $\tilde{X}_{b,i}$ ,  $i=1,2$  en de totalen van de eerste twee kolommen met  $\tilde{X}_{a,j}$  en  $\tilde{X}_{b,j}$ ,  $j=1,2$ , dan kan het minimaliseringsprobleem geschreven worden als:

$$\text{Minimaliseer } D(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{ij} w_{a,ij} (x_{a,ij} - \tilde{x}_{a,ij})^2 + \sum_{ij} w_{b,ij} (x_{b,ij} - \tilde{x}_{b,ij})^2$$

Onder de voorwaarden  $\tilde{X}_{a,i} = \tilde{X}_{b,i}$ ,  $\tilde{X}_{a,j} = \tilde{X}_{b,j}$  voor  $i = 1,2; j = 1,2$

waarbij de index a of b de tabel aanduidt (2a respectievelijk 2b) en de indices i en j lopen over de aan te passen celwaarden. De gewichten ( $w_{a,ij}, w_{b,ij}$ ) geven de mate van betrouwbaarheid van de celwaarden weer (bijvoorbeeld de inverse van de geschatte variantie). Celaantallen met een groot gewicht hebben veel effect op de criterium waarde en zullen daarom minder aangepast worden.

Generalisatie van deze methode naar o.a. algemene lineaire gelijkheidsrestricties en ongelijkheidsrestricties wordt ondermeer besproken in Magnus et al. (2000) en Bikker et al. (2007).

## 2.2 Aanpassen op micro-niveau: micro-integratie

**Tabel 3**  
Een record uit een bedrijfsstatistiek

Variabele	Naam	Survey waarde
$x_1$	Bedrijfsresultaat (Resultaat)	400
$x_2$	Werkzame personen (WP)	30
$x_3$	Omzet totaal (Omzet)	4 400
$x_4$	Omzet hoofdactiviteit (Omzet hoofd)	4 100
$x_5$	Omzet neven activiteiten (Omzet neven)	300
$x_6$	Personeelslasten totaal (Pers. totaal)	2 000
$x_7$	Personeelslasten salarissen (Pers. salaris)	1 500
$x_8$	Personeelslasten overig (Pers. overig)	500
$x_9$	Overige lasten	2 000
$x_{10}$	Totale lasten	4 000

Strategieën voor minimale aanpassingen aan de data zodanig dat de aangepaste data aan bepaalde voorwaarden voldoen zijn ook toe te passen om gegevens uit verschillende bronnen op micro-niveau te integreren. Als voorbeeld beschouwen we het bedrijfsrecord weergegeven in tabel 3.

Dit record moet aan een aantal logische en boekhoudkundige identiteiten voldoen. Dit is te formuleren door de volgende restricties op de variabelen:

$$a_1: x_1 - x_3 + x_{10} = 0 \text{ (Resultaat = Omzet - Totale lasten);}$$

$$a_2: x_3 - x_4 - x_5 = 0 \text{ (Omzet = Omzet hoofd + Omzet neven);}$$

$$a_3: x_6 - x_7 - x_8 = 0 \text{ (Pers. totaal = Pers. salaris + Pers. overig);}$$

$$a_4: x_{10} - x_6 - x_9 = 0 \text{ (Totale lasten = Pers. total + Overige lasten).}$$

In het algemeen is een restrictie  $k$  te formuleren als  $\sum_i a_{ik} \tilde{x}_i = b_k$ , waarbij de sommatie loopt over alle aan te passen variabelen  $x_i$  en  $a_{ik}$  is 0 als variabele  $x_i$  niet in restrictie  $k$  voorkomt en 1 of -1 anderszins.

Veronderstel nu dat er uit betrouwbare administratieve bronnen voor deze respondent de variabelen Omzet en WP beschikbaar zijn en de waarden voor deze varia-

belen verschillen van die uit het survey. Om consistent te blijven met eerdere publicaties van Omzet en WP op basis van de administratieve bronnen zouden we de waarden van de survey-variabelen kunnen vervangen door de administratieve waarden. Het resulterende record zal dan echter niet meer voldoen aan de restricties a1 en a2. Om een consistent aangepast record te krijgen dat de administratieve gegevens bevat voor variabelen waarvoor deze beschikbaar zijn én aan de restricties voldoet kunnen we het volgende minimale aanpasprobleem formuleren:

Minimaliseer  $D(\mathbf{x}, \tilde{\mathbf{x}})$  onder de voorwaarden a1-a4,

waarbij  $\mathbf{x}$  de aan te passen variabelen zijn (hier dus de variabelen waarvoor geen administratieve waarde beschikbaar is) en  $\tilde{\mathbf{x}}$  de aangepaste waarden voor deze variabelen zodanig dat aan de restricties is voldaan. Omdat, althans in dit geval, de waarden van de variabelen die uit de administratieve gegevens afkomstig zijn niet aangepast worden, vullen we die waarden als constanten in de restricties in, zodat de vergelijkingen a1-a4 alleen de variabelen bevatten die voorkomen in  $\mathbf{x}$  (en  $\tilde{\mathbf{x}}$ ) en dus aangepast mogen worden. Behalve de bovenbeschreven gelijkheidsrestricties, zullen we in het vervolg ook ongelijkheidsrestricties beschouwen. Een veel voorkomende ongelijkheidsrestrictie voor economische variabelen is dat de waarden niet negatief mogen zijn. Dit geldt bijvoorbeeld voor alle variabelen uit tabel 3, behalve de eerste. In het algemeen kunnen lineaire ongelijkheidsrestricties geformuleerd worden als  $\sum_i a_{ik} \tilde{x}_i \geq b_k$ .

### 3. *Afstandfuncties en aanpasmodellen.*

Het aanpassen van data zodanig dat aan bepaalde restricties wordt voldaan is in de vorige paragraaf geformuleerd als het minimaliseren van een functie die de afstand of discrepantie weergeeft tussen de oorspronkelijke data en de aangepaste data, onder de voorwaarde dat de aangepaste data aan de restricties moeten voldoen. Dit minimalisatieprobleem, met zowel gelijkheids- als ongelijkheidsrestricties is een standaard optimaliseringsprobleem waarvoor verschillende algoritmen zijn beschreven in de literatuur. Hier gaan we niet in op specifieke algoritmen maar beschrijven we de structuur van de aanpassingen voor verschillende keuzes van de afstandsfunctie. Dit “aanpasmodel” geeft inzicht in de eigenschappen van de aangepaste data en kan behulpzaam zijn bij het kiezen van een aanpasmethode.

Deze structuur kan afgeleid worden door de Lagrangiaan voor het minimalisatieprobleem te beschouwen. Dit is een functie waaruit verschillende eigenschappen van de oplossing kunnen worden afgeleid. Voor de hier beschouwde aanpassingsproblemen is deze functie gegeven door:

$$L(\tilde{\mathbf{x}}, \boldsymbol{\alpha}) = D(\mathbf{x}, \tilde{\mathbf{x}}) + \sum_k \alpha_k \left( \sum_i a_{ik} \tilde{x}_i - b_k \right), \quad (1)$$

met de parameters  $\alpha_k$  de Lagrange multiplicatoren, één voor ieder van de  $K$  res-



tricties en  $\alpha$  de vector met elementen  $\alpha_1, \dots, \alpha_K$ . Eén van de condities die moet gelden voor de oplossing  $\tilde{\mathbf{x}}$  van het minimalisatieprobleem is dat de afgeleide naar  $\tilde{\mathbf{x}}$  van de Lagrangiaan gelijk aan nul moet zijn. Voor de afgeleide  $L'_{x_i}(\tilde{x}_i, \alpha)$  naar een variabele  $\tilde{x}_i$  vinden we

$$L'_{x_i}(\tilde{x}_i, \alpha) = D'_{x_i}(\mathbf{x}_x, \tilde{x}_i) + \sum_k \alpha_k a_{ik} = 0, \quad (2)$$

met  $D'_{x_i}$  de afgeleide van  $D$  naar  $\tilde{x}_i$ . Met behulp van deze conditie is af te leiden hoe de verschillende keuzes voor  $D$  tot verschillende benaderingen voor het aanpasprobleem leiden. Hieronder beschouwen we drie gebruikelijke keuzes voor de afstandsfunctie: kleinste kwadraten (Least Squares, LS), gewogen kleinste kwadraten (Weighted Least Squares, WLS) en de Kullback-Leibler afstand (KL) en tonen de bijbehorende structuur voor de aanpassingen, of het "aanpasmodel".

Voor het kleinste kwadraten criterium,  $\frac{1}{2} \sum_i (\tilde{x}_i - x_i)^2$ , volgt uit (2) dat het aanpasmodel de volgende additieve vorm heeft:

$$\tilde{x}_i = x_i + \sum_k a_{ik} \alpha_k. \quad (3)$$

Dit laat zien dat volgens het criterium LS de totale aanpassing voor een variabele  $\tilde{x}_i$  de som is van aanpassingen voor ieder van de  $K$  restricties. De aanpassingsparameter  $\alpha_k$  beschrijft de mate van aanpassing voor restrictie  $k$  en verschilt voor de restricties maar is gelijk voor alle variabelen. De variabelen  $a_{ik}$  zijn 0, 1 of -1 en geven aan of er, als gevolg van restrictie  $k$ , geen aanpassing plaats vindt (als  $\tilde{x}_i$  niet voorkomt in restrictie  $k$ ) of een toename of een afname.

Voor het gewogen kleinste kwadraten criterium,  $\frac{1}{2} \sum_i w_i (\tilde{x}_i - x_i)^2$ , volgt uit (2) de volgende gewogen additieve structuur voor de aanpassingen:

$$\tilde{x}_i = x_i + \frac{1}{w_i} \sum_k a_{ik} \alpha_k. \quad (4)$$

In tegenstelling tot het ongewogen kleinste kwadraten criterium geldt voor WLS dat de mate van aanpassing aan een restrictie  $k$  niet gelijk is voor alle variabelen, de mate van aanpassing varieert per variabele volgens de gewichten  $w_i$ ; variabelen met grote gewichten hebben meer invloed op het te minimaliseren criterium en worden daarom minder aangepast dan variabelen met kleine gewichten.

Voor het criterium KL,  $\sum_i \tilde{x}_i (\ln \tilde{x}_i - \ln x_i)$ , kan aangetoond worden dat de aanpassingen de volgende multiplicatieve vorm hebben:

$$\tilde{x}_i = x_i \times \prod_k \exp(-a_{ik} \alpha_k). \quad (5)$$

De totale aanpassing voor een variabele  $\tilde{x}_i$  is het product van aanpassingen aan ieder van de restricties. De factor  $\exp(-a_{ik}\alpha_k)$  in dit product geeft de aanpassing weer voor restrictie  $k$  en is 1 voor  $a_{ik}$  is 0 (geen aanpassing),  $1/\exp(\alpha_k)$  als  $a_{ik}$  is 1 en de inverse van deze factor, als  $a_{ik}$  is -1.

#### 4. *Verschillende opties voor het aanpassen van het bedrijfsrecord.*

In deze paragraaf zullen we een aantal verschillende mogelijkheden voor het aanpassen van een bedrijfsrecord, zoals getoond in tabel 3, aan de hand van een eenvoudig voorbeeld illustreren. We beschouwen de verschillende aanpasmodellen, corresponderend met de verschillende afstandsfuncties, het aanpassen van slechts een deel van de variabelen die kunnen worden aangepast, het opleggen van bepaalde verhoudingen aan de aangepaste variabelen, het opleggen van de restrictie dat een survey variabele weliswaar gelijk moet zijn aan de administratieve variabele maar dat beide variabelen mogen worden aangepast in plaats van alleen de survey variabele en, ten slotte, het aanpassen van geïmputeerde waarden.

**Tabel 4**  
Aanpassen van de data uit tabel 3 aan een administratieve waarde voor Omzet met behulp van verschillende aanpasmodellen

Var.	Naam	Survey waarde	Aangepaste waarde		
			LS	LS niet-negatief	WLS/KL
$x_1$	Resultaat	400	-30,8	-30,8	267,6
$x_2$	WP	30	25	25	25
$x_3$	Omzet	4 400	3 700	3 700	3 700
$x_4$	Omzet hoofd	4 100	3 750	3 700	3 447,7
$x_5$	Omzet overig	300	-50	0	252,3
$x_6$	Pers. totaal	2 000	1 892,3	1 892,3	1 810,8
$x_7$	Pers. salaris	1 500	1 446,2	1 446,2	1 358,1
$x_8$	Pers. overig	500	446,2	446,2	452,7
$x_9$	Overige lasten	2 000	1 838,5	1 838,5	1 621,6
$x_{10}$	Totale lasten	4 000	3 730,8	3 730,8	3 432,4

#### *Consistentie met behulp van verschillende aanpasmodellen.*

In dit voorbeeld veronderstellen we dat er voor de variabelen WP en Omzet uit het bedrijfsrecord uit tabel 3 gegevens uit betrouwbare administratieve bronnen beschikbaar zijn en dat de administratieve waarde voor WP gelijk is aan 25 en voor Omzet gelijk is aan 3700. Als we deze waarden overnemen in het record worden de restricties waarin Omzet voorkomt geschonden. Voor WP geldt dat niet omdat WP in geen enkele restrictie voorkomt. Om het record consistent te maken met de restricties zijn de verschillende aanpasmodellen toegepast, de resultaten staan in tabel 4. De kolom "LS" geeft de aangepaste waarden volgens het kleinste kwadraten criteri-

um. De omzet in deze kolom is gefixeerd op de administratieve waarde, de andere variabelen zijn aangepast (zo weinig mogelijk) om aan de restricties te voldoen. Een gevolg van deze aanpassingen is dat de variabele Resultaat vrij sterk veranderd en van positief naar negatief is gegaan. Een ander gevolg is dat Omzet overig negatief is geworden wat geen toelaatbare waarde is voor deze variabele. Om dat laatste probleem op te lossen is als extra restrictie toegevoegd dat Omzet overig groter of gelijk aan nul moet zijn wat leidt tot de resultaten in de kolom LS niet negatief. Deze resultaten zijn identiek aan de resultaten in de kolom LS behalve dat Omzet overig nu gelijk aan nul is geworden waardoor Omzet hoofd gelijk moet worden aan 3 700. Omdat sommige vrij kleine bedragen relatief veel lijken te worden aangepast is het aanpasmodel volgens het gewogen kleinste kwadraten criterium toegepast met als gewichten de inverse van de oorspronkelijke waarden, variabelen met een grote oorspronkelijke waarde krijgen hierdoor een klein gewicht en zullen, in absolute zin, meer aangepast worden dan variabelen met een kleine oorspronkelijke waarde. Ook is het multiplicatieve aanpasmodel volgens het KL criterium toegepast. Deze beide methoden leverden echter identieke resultaten op<sup>3</sup> en deze zijn weergegeven in de kolom WLS/KL. Uit deze resultaten blijkt duidelijk dat de aanpassingen voor de kleine bedragen veel minder groot zijn. De waarde voor Resultaat ligt veel dichterbij de oorspronkelijke waarde en blijft dan ook positief en ook Omzet overig wordt veel minder aangepast en blijft nu positief zonder extra ongelijkheidsrestrictie te hoeven invoeren. Het verschil tussen LS en WLS/KL komt duidelijk naar voren bij  $x_7$  en  $x_8$  omdat deze variabelen slechts in één restrictie voorkomen. De aanpassing is even groot voor deze twee variabelen bij LS en proportioneel bij WLS/KL. Een gevolg hiervan is dat bij WLS/KL de oorspronkelijke verhouding tussen  $x_7$  en  $x_8$  behouden blijft terwijl dit voor LS niet het geval is. Bij alle volgende alternatieven is steeds het WLS criterium gebruikt.

*Consistentie met administratieve Omzet, zonder alle variabelen aan te passen.*

**Tabel 5**  
Aanpassen met het WLS-model onder alternatieve assumpties

Var.	Naam	Survey/register waarde	Aangepaste waarde		
			$x_6$ - $x_8$ niet aanpassen	$x_7 = 50 \times x_2$	Gelijkheid omzet
$x_1$	Resultaat	400	238,5	278,9	349,7
$x_2$	WP	30	25	25	25
$x_3$	Omzet	4 400	3 700	3 700	4 134,2
$x_4$	Omzet hoofd	4 100	3 447,7	3 447,7	3 831,7
$x_5$	Omzet overig	300	252,3	252,3	302,5
$x_6$	Pers. Totaal	2 000	2 000	1 736,8	1 928,2
$x_7$	Pers. salaris	1 500	1 500	1 250	1 446,1
$x_8$	Pers. overig	500	500	486,8	482
$x_9$	Overige lasten	2 000	1 461,5	1 684,2	1 856,3
$x_{10}$	Totale lasten	4 000	3 461,5	3 421	3 784,5

De voorgaande oplossingen gingen er van uit dat alle variabelen die aangepast kunnen worden ook aangepast worden. In de praktijk is echter gebleken dat een verandering in Omzet niet altijd sterk correleert met een verandering in de personeelslasten (althans op korte termijn). Als we de data willen aanpassen om te voldoen aan een andere waarde voor Omzet, kan er voor gekozen worden om de variabelen  $x_6$ - $x_8$  niet aan te passen. Ook zonder die waarden aan te passen kan een consistent record worden gecreëerd, dat aan alle vereiste restricties voldoet. In tabel 5 staan in de kolom " $x_6$ - $x_8$  niet aanpassen" de resultaten voor deze benadering. De aangepaste omzet variabelen zijn identiek aan die in de laatste kolom van tabel 4, omdat de waarden van de personeelslasten hier niet op van invloed zijn. De andere aangepaste variabelen verschillen niet heel veel van die in de laatste kolom van tabel 4. In zekere zin is het aanpassen van minder variabelen minder optimaal dan het aanpassen van alle daarvoor in aanmerking komende variabelen. Het WLS criterium voor het aanpassen van alle mogelijke variabelen (de oplossing in de laatste kolom van tabel 4) heeft de waarde 343,1 terwijl dit criterium de waarde 394,1 heeft voor het aanpassen van minder variabelen. Als de veronderstelling van een slechts zwakke relatie tussen Omzet en de personeelslasten juist is, is de keuze voor het niet aanpassen van de personeelslasten te rechtvaardigen en ontstaat ondanks de grotere waarde van het WLS criterium toch een "beter" record.

#### *Consistentie met administratieve Omzet en WP en een bekende verhouding.*

Tot nu toe is er geen gebruik gemaakt van de administratieve variabele WP. Net als de omzet is ook deze waarde lager dan de waarde van de overeenkomstige survey variabele. De variabele WP komt echter in geen van de restricties voor en daarom heeft de nieuwe waarde van deze variabele geen invloed op de te veranderen variabelen in het record. Als we er van uit gaan dat de administratieve waarden de "goede" waarden zijn en de verhouding tussen WP en Pers. salaris gelijk moet zijn aan wat in de oorspronkelijke survey data was geobserveerd, dan gaat hiermee informatie verloren. Deze informatie is te gebruiken door de restrictie toe te voegen dat Pers. salaris =  $50 \times WP$ . Deze extra restrictie resulteert in de aangepaste waarden in de kolom " $x_7 = 50 \times x_2$ " in tabel 5. De aangepaste waarde van Pers. salaris is nu duidelijk lager dan zonder deze restrictie en in overeenstemming met het lagere aantal WP. De verhouding tussen Pers. salaris en Pers. overig is nu echter, in tegenstelling tot de eerdere WLS oplossingen, niet meer gelijk aan de oorspronkelijke 3:1 verhouding omdat Pers. salaris nu voorkomt in twee restricties en aan beiden aangepast moet worden. Desgewenst is voor deze verhouding ook weer de oorspronkelijke verhouding te reproduceren door een extra restrictie toe te voegen. In de hier beschreven oplossing wordt de verhouding tussen WP en Pers. salaris uit de oorspronkelijke data exact gereproduceerd omdat deze verhouding expliciet als restrictie, Pers. salaris =  $50 \times WP$ , is meegenomen. Het is ook mogelijk om verhoudingen en andere a priori kennis als "zachte" restricties mee te nemen. De data worden dan zodanig aangepast dat eventuele schendingen van zachte restricties

weliswaar verminderd worden maar niet noodzakelijk geheel opgelost worden (zie Magnus et al. (2000) en Bikker et al. (2007))

### *Gelijkheid tussen Omzet uit verschillende bronnen.*

Als laatste benadering voor het aanpasprobleem verlaten we het uitgangspunt dat in het aangepaste record de administratieve waarden exact gereproduceerd moeten worden. We vervangen nu niet de survey variabele Omzet door de administratieve waarde van deze variabele (die dan vervolgens als vast en niet aanpasbaar wordt beschouwd) maar we nemen twee omzetvariabelen in het aan te passen record mee, de administratieve versie en de survey versie. Deze beide variabelen mogen aangepast worden, maar we leggen wel de extra restrictie op dat ze aan elkaar gelijk moeten zijn, zodat er toch maar één unieke waarde voor omzet in het uiteindelijke record voor komt. De resultaten van deze exercitie staan in de laatste kolom van tabel 5. De aangepaste waarde van omzet is nu niet meer gelijk aan de administratieve waarde 3 700, maar komt tussen deze waarde en de oorspronkelijke 4 400 in te liggen. Ook de ander variabelen komen nu dichterbij de oorspronkelijke waarden te liggen. Als we twee bronnen hebben waarvan we niet één willen aanmerken als de referentie bron (omdat deze betrouwbaar geacht wordt of omdat we consistentie met eerder gepubliceerde cijfers willen bereiken), dan is een methode waarbij de uiteindelijke waarde ergens tussen de beide bronnen in komt te liggen een optimale keuze. Met behulp van gewichten die de betrouwbaarheid van beide bronnen weergeven is dan in te stellen dat we dichterbij de meest betrouwbare bron willen uitkomen.

### *Het aanpassen van geïmputeerde waarden.*

Regelmatig komt het voor dat de beoogde respondenten niet antwoorden (non-respons) of niet alle vragen beantwoorden (partiële non-respons). Eén van de methoden om voor (partiële) non-respons te corrigeren is imputatie: het invullen van een voor spelde waarde, volgens een model, voor de ontbrekende waarde. Een vorm van imputatie die door statistische bureaus veel wordt toegepast bij economische statistieken is ratio-imputatie. Veronderstel dat in het voorbeeld record de personeelsgegevens WP, Pers. totaal, Pers. salaris en Pers. overig ontbreken maar dat er uit een administratieve bron wel een waarde voor WP beschikbaar is. De ontbrekende waarde voor Pers. totaal kan dan geïmputeerd worden door de ratio Pers. totaal/WP te schatten en de administratieve waarde van WP met deze ratio te vermenigvuldigen<sup>4</sup>. De variabelen Pers. salaris en Pers. overig kunnen vervolgens geïmputeerd worden door de geschatte ratio's van deze variabelen tot de geïmputeerde Pers. totaal te gebruiken. In tabel 6 wordt dit geïllustreerd. De administratieve waarde van WP is 25. Uit een vorige opgave van dit bedrijf bleek de verhouding Pers. totaal/WP gelijk te zijn aan  $2000/30=66,67$ . Vermenigvuldigen met 25 levert de geïmputeerde waarde  $1\ 666,67$  voor Pers. totaal. De variabelen Pers. salaris en Pers. overig worden geïmputeerd door gebruik te maken van de geschatte ratio's van deze variabelen tot Pers. totaal (de aandelen).

Tabel 6  
Aanpassen van geïmputeerde waarden

	Pers. Totaal	Pers. salaris	Pers. overig
Aandeel van <i>Pers. Totaal</i>	1	0,75	0,25
Geïmputeerde waarde	1 666,67	1 250	416,67
Aangepaste geïmputeerde waarde	1 753,55	1 315,17	438,39

Een record met deze imputaties zal echter niet meer aan alle restricties voldoen, voor de geïmputeerde waarde voor Pers. totaal hoeft niet te gelden dat Totale lasten = Pers. totaal + Overige lasten. Het geïmputeerde record kan daarom aangepast worden zodanig dat aan alle restricties wordt voldaan. Als hiervoor het multiplicatieve model gebruikt wordt krijgen we de resultaten in de laatste rij van tabel 6. Door de multiplicatieve structuur van deze aanpassingen blijven de verhoudingen 0,75 en 0,25 behouden. De verhouding tussen de vaste waarde van WP en Pers. totaal is echter veranderd omdat zonder aanpassing van Pers. totaal niet aan bovengenoemde restrictie kan worden voldaan.

## 5. Conclusie

De optimaliseringsmethoden die ontwikkeld zijn voor integratieproblemen op macro-niveau kunnen ook toegepast worden om op micro-niveau informatie uit verschillende data bronnen, “harde” kennis (zoals boekhoudkundige en logische restricties) en “zachte” kennis (zoals verwachtingen over verhoudingen) te integreren. De optimalisatiemethoden verschillen in de doelfunctie die geminimaliseerd wordt, maar in dit rapport is benadrukt dat zij verschillende modellen impliceren voor de aanpassingen aan de data die nodig zijn om aan alle voorwaarden te voldoen. Het resultaat van deze micro-integratie stap hangt af van het gespecificeerde model (een (gewogen) additieve of multiplicatieve structuur van de aanpassingen), de keuze welke variabelen wel of niet aangepast mogen worden, de instelling van betrouwbaarheidsgewichten en de specificatie van de harde en zachte kennisregels. Als de inconsistenties tussen de verschillende databronnen en de harde kennisregels klein zijn, zijn de bovenstaande keuzes vaak van minder belang omdat de aanpassingen dan gering zullen zijn, al kunnen vele kleine aanpassingen schattingen van totalen of verbanden tussen variabelen (zoals verhoudingen) verstoren als het aanpasmodel daar geen rekening mee houdt. Als er grote inconsistenties tussen databronnen voorkomen en de definities van variabelen en meetmomenten zijn in overeenstemming, dan komen er waarschijnlijk grote fouten in de data voor. Deze fouten zouden eerst geïdentificeerd en zo mogelijk gecorrigeerd moeten worden voordat er een automatisch aanpassingsalgoritme wordt toegepast. Inconsistenties kunnen ook ontstaan doordat sommige ontbrekende waarden met een model zijn geïmputeerd of omdat de meetmomenten van verschillende bronnen niet

overeenkomen. In zulke gevallen zal het gekozen aanpasmodel en de daarbij gebruikte kennisregels zorgvuldig gespecificeerd moeten worden omdat het aanpassen dan van invloed kan zijn op de eigenschappen van de uitkomsten. In het geval van imputatie moet het aanpasmodel gezien worden als onderdeel van het imputatiemodel en in het geval van verschillende waarnemingsmomenten is het aanpasmodel onderdeel van de methode voor extrapolatie van oudere waarnemingen naar een recenter tijdstip.

## Referenties

Bikker, R., J. Daalmans en N. Mushkudiani (2007). *Methodenreeks: thema Macro-integratie, deelthema inpassen*. Rapport, Centraal Bureau voor de Statistiek, Den Haag.

Bishop, Y.M.M., S.E. Fienberg and P.W. Holland (1975), *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press, Cambridge, Massachusetts.

Deming, W. E., en F.F. Stephan (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*. 11, pp. 427–444.

De Waal, T., J. Pannekoek en S. Scholtus (2011). *Handbook of Statistical Data Editing and Imputation*. Wiley, New York.

Friedlander, D., 1961, A technique for estimating a contingency table given the marginal totals and some supplementary data. *Journal of the Royal Statistical Society, Series A*, pp. 412–420.

Magnus, J.R., J.W. van Tongeren en A.F. de Vos (2000), National Accounts Estimation Using Indicator Ratios, *The Review of Income and Wealth* 3, 329–350.

Stone, J.R.N., D.A. Champerowne and J.E. Maede (1942), The Precision of National Income Accounting Estimates. *Reviews of Economic Studies* 9, pp. 111–125.

## Noten

- 
- <sup>1.)</sup> Dit voorbeeld is beschreven in Bishop et al. (1975, pag. 98) waaruit deze tabel (enigszins gecomprimeerd) is overgenomen.
  - <sup>2.)</sup> In werkelijkheid zijn aanbod- en gebruiktabellen veel ingewikkelder. Niet alleen omdat er veel meer bedrijfstakken en goederengroepen worden onderscheiden maar ook omdat er rekening gehouden wordt met import, export, productgebonden belastingen en subsidies en handels- en vervoersmarges.
  - <sup>3.)</sup> In het algemeen kan aangetoond worden dat de criteria WLS en KL bij benadering tot dezelfde resultaten leiden als de gekozen gewichten voor WLS gelijk zijn aan de inverse van de oorspronkelijke waarden, voor andere gewichten geldt dat niet.
  - <sup>4.)</sup> Voor grote bedrijven die jaarlijks in de steekproef zitten kan deze ratio geschat worden met een vorige opgave van hetzelfde bedrijf. Als er geen historische gegevens beschikbaar zijn kan de ratio geschat worden met behulp van de gegevens van soortgelijke responderende bedrijven.



# *Discussie over Pannekoek, J.: “Modellen voor micro-integratie”*

Stef van Buuren (TNO/UU)

De bijdrage van Pannekoek geeft een overzicht van technieken voor het combineren van data uit verschillende bronnen. De gepresenteerde methoden generaliseren het krachtige Iterative Proportional Fitting (IPF) algoritme naar optimalisatie onder voorwaarden. Het inbouwen van meta-kennis over de data zorgt voor betere schattingen. Pannekoek (2010) demonstreert op overtuigende wijze hoe dergelijke generalisaties waardevol zijn voor de praktijk. In deze notitie vraag ik aandacht voor enkele technische punten.

De modellen in de bijdrage richten zich op data fitting. De tekst specificeert niet expliciet voor welke populatie de schattingen geldig moeten zijn, bijvoorbeeld voor alle inwoners van Nederlands, of voor alle Nederlandse bedrijven met meer dan 50 werknemers. Zonder zo'n expliciete definitie is het niet goed mogelijk te bepalen of de gehanteerde techniek ook daadwerkelijk leidt tot de juiste antwoorden.

De tekst is opgebouwd rondom het probleem hoe te komen tot schattingen voor het binnenwerk van een tabel wanneer de marginalen bekend zijn. Er wordt weinig gezegd over waar deze marginalen vandaan komen. Naar mijn idee bepaalt de herkomst van de marginalen mede hoe je de schattingstechniek zal inrichten en gebruiken. Ik denk daarom dat het nuttig is onderscheid te maken tussen een aantal situaties:

In het meest eenvoudige geval zijn de marginalen afkomstig uit eenzelfde registratie. Het is dan in principe ook mogelijk om direct het binnenwerk te berekenen. De gepresenteerde technieken zijn waarschijnlijk niet echt voor dit geval bedoeld, maar deze situatie is wel nuttig voor validatie. We kunnen immers de schattingen vergelijken met de daadwerkelijke geobserveerde tellingen.

Indien de marginalen uit verschillende registraties komen, dan worden we geconfronteerd met de vraag of de populatiedekking van beide registraties voldoende overeenkomt. Hiermee hangt samen dat de geobserveerde totaal aantallen kunnen verschillen, waardoor het onmogelijk wordt om een binnenwerk te maken waarvoor de rij- en kolomtotalen tot hetzelfde getal optellen. Het is niet duidelijk wat er in beide gevallen gebeurt. Het zou nuttig zijn beide complicaties expliciet te behandelen.

Een situatie die op II. lijkt, maar toch wezenlijk anders is krijgen we wanneer de marginalen worden geschat op basis van (verschillende) steekproeven. De margi-

nal en dan schattingen, en kennen zelf een onzekerheid. Indien je deze schattingen behandelt als ware zij het werkelijk geobserveerde tellingen, dat past het binnenwerk zich aan aan de toevallige steekproeffluctuaties in beide databronnen. Bij mijn weten bestaat er geen formele statistische theorie over dit geval. Mijn intuïtie zegt me dat naarmate je meer bronnen integreert je een steeds groter risico loopt dat je logisch onmogelijke combinaties moet maken. Daarnaast denk ik dat inconsistentie in de marginalen ernstiger wordt naarmate de steekproefomvang in de gebruikte bronnen meer uiteenlopen. Een mogelijke oplossing voor dit geval zou kunnen zijn: maak een random trekking uit beide marginale verdelingen, en schat vervolgens het binnenwerk. Herhaal deze twee stappen 100 keer, en middel de 100 schattingen van het binnenwerk. Ik verwacht dat dit een ander resultaat zal opleveren dan negeren van de onzekerheid, en dat het verschil groeit naarmate de steekproefomvang van de databronnen uiteenlopen.

Het Expectation-Conditional Maximization (ECM) algoritme (Meng en Rubin, 1993) is een generalisatie van het EM algoritme. Meng en Rubin laten zien dat het ECM algoritme ook een generalisatie is van IPF. Een specifiek voordeel van ECM algoritme binnen de huidige context is dat het een theoretisch goed gefundeerde wijze biedt om ontbrekende gegevens te behandelen. Het is ook mogelijk om de verdeling van het binnenwerk zelf te simuleren gegeven de marginalen. Methoden hiervoor staan bekend als Bayesian IPF, en zijn voorgesteld door Gelman et al. (1995) en Schafer (1997, H.8).

Pannekoek geeft de volgende omschrijving van imputatie: "het invullen van een voorspelde waarde, volgens een model, voor de ontbrekende waarde". Veel lezers zullen dit opvatten dat de geïmputeerde waarde gelijk wordt gesteld aan de voorspelde (en meest waarschijnlijke) waarde. Dat is niet altijd een gelukkige keuze. Wanneer we per respondent de verwachte waarde uit het model invullen, dan leidt dat tot de kunstmatige reductie van de spreiding van de compleet gemaakte variabele. De reductie is sterker naarmate de verklaarde variantie van het imputatiemodel afneemt. Tevens leidt de methode tot een overschatting van de correlatie van de aangevulde variabele met de predictor variabelen. Beide problemen zijn op te heffen door een random trekking uit de residuen aan de voorspelde waarde toe te voegen. Daarnaast dient ook de onzekerheid van het imputatiemodel zelf in de imputatie te worden opgenomen. Aangezien het CBS vaak met grote bestanden werkt is de invloed van de laatste mogelijk gering. Geïmputeerde waarden op basis van alleen de verwachte waarde hebben een beperkt gebruiksgebied. Aggregeren over deze imputatie leidt tot foutieve resultaten. De gebruiker is zich daarvan niet altijd bewust.

Ik sluit af met de constatering dat de bijdrage van Pannekoek een welkome aanvulling is op het arsenaal van methoden voor het schatten van het binnenwerk van de tabellen. Hopelijk draagt mijn commentaar bij aan een verdere verfijning van deze methoden.

## *Referenties*

Gelman, A., Rubin, D.B., Carlin, J., Stern, H. (1995). *Bayesian Data Analysis*. Chapman & Hall, London.

Meng, X.L., Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80(2):267–278.

Pannekoek, J. (2010). Modellen voor micro-integratie. Bijdrage gepresenteerd op het Congres “Registers in sociaalwetenschappelijk onderzoek: mogelijkheden en valkuilen”, 24 juni, Naturalis, Leiden.

Schafer J.L. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall, London.



# *Registers: Onderzoeksagenda voor de toekomst*

Piet J. H. Daas, Léander Kuijvenhoven en Kees Zeelenberg (CBS)

## *1 Inleiding*

De combinatie van de toenemende maatschappelijke en politieke behoefte aan informatie, de afname in de respons op enquêtes en de toename in de beschikbaarheid van registers heeft tot gevolg dat sociaalwetenschappelijk, medisch en statistisch onderzoek steeds vaker gebaseerd zijn op registergegevens (Hox en Boeije, 2005; Bethlehem, 2008; Bakker, 2009). Het gebruik van registergegevens heeft dan ook een aantal voordelen. Zo zijn de gegevens al verzameld, in het algemeen voor een groot deel van de populatie beschikbaar en beschrijven ze, in het ideale geval, meerdere aansluitende perioden. Ook het Centraal Bureau voor de Statistiek (CBS) is zich hier terdege bewust van. Steeds vaker zijn sociale en ruimtelijke statistieken op registergegevens gebaseerd (Bakker, 2009). Het gebruik van registers heeft echter ook nadelen. Zo is de toegang soms lastig, is de controle op de kwaliteit beperkt, zijn de gegevens meestal voor een ander doel verzameld dan het doel waarvoor de onderzoeker ze zou willen gebruiken en is de set van bruikbare variabelen beperkt (Wallgren en Wallgren, 2007). Doordat het CBS het wettelijke recht heeft gekregen van kosteloze toegang tot veel registers is het eerste nadeel vrijwel opgelost. De laatste twee nadelen kunnen aanzienlijk verminderd worden door het grootschalig combineren van registerbestanden. De gecombineerde bestanden bieden tevens nieuwe onderzoeksmogelijkheden (Bethlehem, 2008; Bakker 2009). Die mogelijkheden en de valkuilen die daarbij op kunnen treden bepalen de onderzoeksagenda voor de toekomst.

De artikelen in deze bundel geven een goed beeld van het onderzoeksveld van registers en daarmee ook van de uitdagingen op dit terrein. Dit afsluitende artikel zal voornamelijk gaan over de methodologische vragen en wat minder over analyse gemotiveerd vanuit de inhoud. Ook de interactie en noodzaak van samenwerking tussen beide komt kort naar voren. Het bespreken van de methodologische uitdagingen is gebaseerd op de structuur van het verwerkingsproces van registers op het CBS. Uiteindelijk doel van dit proces is het produceren van tabellen en microdata van hoge kwaliteit.

Essentieel bij het gebruik van registergegevens is de kwaliteitsvraag. Wat is de kwaliteit, en hoe bepaal je de kwaliteit van registers en microdata? Dit is een belangrijk onderwerp. Registers maar ook enquêtes bevatten nu eenmaal fouten. Zo verscheen twee jaar geleden een bericht dat in de bevolkingsboekhouding, de Ge-

meentelijke basisadministratie (GBA), van Amsterdam bijna 15% van de records fouten bevat (Gemeente Amsterdam, 2008). Ook in deze bundel is kwaliteit een aantal maal aan de orde geweest. De oorzaken van slechte kwaliteit zijn velerlei. Ze kunnen variëren van kostenoverwegingen bij de registratiehouder tot typefouten of fraude. We kunnen vaak niet zoveel doen aan die oorzaken. Maar van de gevolgen, meetfouten, hebben we last omdat ze kunnen leiden tot fouten in analyses en in tabellen.

Vooraf enkele uitgangspunten, stellingen, die een rol spelen bij de bespreking van de onderzoeksvragen van de toekomst:

- Het statistische proces is er op gericht om de kwaliteit van de data en de gebruiksmogelijkheden te verhogen.
- De sociale wetenschapper en de beleidsanalyse zijn beter af met de bewerkte microdata dan met de ruwe data.
- De methodologische vragen bij dat statistische proces leveren bijzonder interessante onderzoeksonderwerpen, juist ook voor sociale wetenschappers.

Het goed zou zijn als sociale wetenschappers meedoen aan de beantwoording van die methodologische vragen. Dat gebeurt al, maar er zijn ook onderzoeksvragen en -velden waar meer samenwerking niet alleen goed zou zijn maar ook noodzakelijk. Dat betekent niet dat sociale wetenschappers procesmethodologen zouden moeten worden. Zoals we hieronder zullen zien, kunnen juist hun inhoudelijke kennis en hun analyses van groot belang zijn.

## 2. *Registers en de toekomst*

De methodologische uitdagingen op registergebied zullen worden besproken aan de hand van de structuur van het statistische verwerkingsproces (Willeboordse, 2008), zoals weergegeven in Figuur 1. Dit leidt uiteindelijk tot een overzicht van de belangrijke methodologische onderzoeksvragen op registergebied.

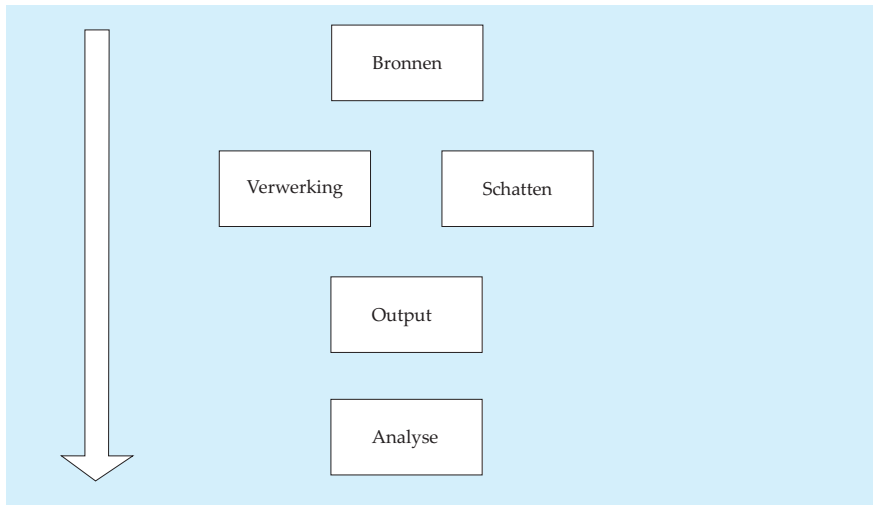
### 2.1 *Het statistische proces*

De kern van het statistische verwerkingsproces, begint met de ontvangst van de bron (de input) en eindigt met de publicatie van de betreffende statistiek of het betreffende microdatabestand (de output). Daarna vindt de analyse plaats. Figuur 1 geeft het verwerkingsproces en de belangrijkste stappen weer. Een onderzoeker die gebruik maakt van registers voor zijn/haar onderzoek zal deze op een vergelijkbare wijze verwerken; alleen wat kleinschaliger. Elk van de stappen in het verwerkingsproces lopen we kort langs.

#### *Bronnen*

De te gebruiken bronnen bevatten alle microdata en zijn hetzij registers en andere administratieve bronnen, hetzij enquêtegegevens, hetzij een combinatie. Een vraag

Figuur 1. Overzicht van het statistisch proces



die zich hierbij opdringt is: moet het CBS alle mogelijke registers binnenhalen, of alleen die waar nu behoefte aan is? Deze vraag is lastig te beantwoorden, maar het is belangrijk hier te constateren dat het CBS alleen gegevens dient te verzamelen die het ook daadwerkelijk gebruikt.

### *Verwerking*

Het proces tussen de ontvangst van de bron en de creatie van de output bestaat uit een systematische reeks bewerkingen op de bron, samengevat onder de noemer verwerking. Deze stap bestaat uit het corrigeren, complementeren en integreren van de data. Waar nodig worden fouten in de data gecorrigeerd, het zogenaamde gaafmaken, dat bestaat uit het opsporen en het corrigeren van die fouten. Ook worden ontbrekende (missende) gegevens soms geïmputeerd. Daarna worden gegevens in verschillende bronnen gekoppeld en worden gegevens geïntegreerd (Willeboordse, 2008). Dit kan op micro- en op macroniveau plaatsvinden. Micro-integratie is het consistent maken van gegevens uit meerdere bronnen voor één bedrijf of één huishouden. Soms worden ook tabellen uit meerdere bronnen consistent gemaakt; dit is macro-integratie.

Alle verwerkingsstappen zijn gericht op kwaliteitsverbetering en vergroting van de mogelijkheden voor analyse. Wat dat laatste betreft: maatschappelijke verschijnselen zijn naar hun aard complex en analyse ervan vereist samenhangende informatie uit meerdere bronnen. Daarom zijn koppelen en integratie belangrijke onderdelen van het statistische proces.

We zien hier de eerste stelling terugkomen, namelijk dat het statistische proces erop gericht is om de kwaliteit van de data en de gebruiksmogelijkheden te verhogen.

### *Schatten*

Het onderdeel schatten betreft het creëren van de output, het bepalen van weegfactoren voor microdatabestanden, het maken van tabellen en natuurlijk ook het schatten voor kleine gebieden. Wanneer veel bestanden met elkaar gecombineerd worden is het belangrijk de 'juiste' methode te kiezen voor het maken van schattingen.

### *Output*

Belangrijkste vraag bij deze stap is of het mogelijk is om met de beschikbare registers de gewenste statistieken en microdatabestanden (de output) te verkrijgen? Voor statistieken kunnen enquêtes als aanvullende bron van gegevens gebruikt worden. Bijvoorbeeld voor het bepalen van nadere detaillering (structuurinformatie), voor het opsplitsen in deelposten, of voor controle van de resultaten uit de registers (benchmarking). Ook de oorzaken en gevolgen van fouten in de registers en het verwerkingsproces beïnvloeden de resultaten van de output.

### *Analyse*

Analyse is uiteraard voor en soms door de eindgebruiker. Voor een klein deel gebeurt dat door het CBS als duiding van de cijfers. Maar belangrijker is uiteraard de analyse door u, de wetenschappelijke onderzoekers, van beleidsproblemen of uit wetenschappelijke interesse. Daar hebben we vandaag allerlei voorbeelden van gezien op het gebied van de economie, arbeidsmarkt, sociale dynamiek en zorg. Maar er zijn er uiteraard nog vele andere.

Onze stelling hier is: hoe beter de data, hoe beter de analyse. In de woorden van de stelling aan het begin: de sociale wetenschapper en de beleidsanalyse zijn beter af met de bewerkte microdata dan met de ruwe data. Uiteindelijk bepaalt zo het statistische proces voor een niet onbelangrijk deel de uitkomsten van onderzoek!

## **2.2 Methodologische onderzoeksvragen**

We zijn nu bij de onderzoeksvragen aanbeland, die voortvloeien uit het statistisch gebruik van registers en uit de inhoudelijke analyse. De 6 belangrijkste gebieden die we hierbij kunnen onderscheiden, zijn: Kwaliteit en validiteit, Gaafmaken, Combineren van bronnen, Schatten, Datavisualisatie en Analyse.

### **2.2.1 Kwaliteit en validiteit**

Het eerste onderwerp is kwaliteit en validiteit van de data zelf: hoe goed zijn de gegevens in het register of het microdatabestand voor het doel van de onderzoeker.

Belangrijk is om dit te bestuderen voor zowel de bron als voor de output. Hiervoor zijn kwaliteitskaders ontwikkeld. Voor de bron, de kwaliteit van de input, is door het CBS een kader opgesteld (Daas et al., 2010). Dat de aanpak werkt laten de resultaten van de toepassing op een aantal registers zien. Belangrijk is dat de kwaliteit van het registers zo snel mogelijk na ontvangst bepaald kan worden. Mogelijk bieden visua-



lisatietechnieken hiervoor een oplossing. Naast het meten van de inputkwaliteit speelt de vraag mee wat een onderzoeker moet doen als de kwaliteit van het register niet goed of goed genoeg is? Kan een onderzoeker of het CBS eisen opleggen aan registerhouders? Is het mogelijk om afspraken te maken met de leverancier van een register over het leveren van kwaliteitsindicatoren? Indien hierop bevestigend geantwoord kan worden, moet onderzocht worden welke indicatoren hiervoor dan het meest informatief zijn en tegelijkertijd de minste inspanning bij de registerhouder leggen.

Voor de output bestaan al langer internationale kaders, bijvoorbeeld Eurostat (2009). Het kader van Eurostat is echter nog niet volledig op registers ingesteld.

Het zou goed zijn als bron en output steeds beschreven worden door middel van die kaders, zodat de beschrijvingen vergelijkbaar zijn.

Kwaliteit was ook één van de onderwerpen die in het artikel van Catrien Bijleveld (2010) besproken werden. Zij liet zien dat de registers op het terrein van criminaliteit serieuze kwaliteitsproblemen kennen. Zo wordt maar 5% van de delicten ermee gedekt. Dit komt vanwege geringe aangifte en nog geringere opheldering. Interessant is ook haar bewering dat als een deviante categorie personen wordt onderzocht, in dit geval een groep mannen die in hun jeugd behandeld waren voor ernstige gedragsproblemen, er in registraties voor zo'n groep vaak veel minder goede gegevens worden aangetroffen.

Bart Bakker (2010) stelde in zijn artikel validiteit aan de orde, d.w.z. de vraag: meten de begrippen in de bron waar de onderzoeker naar op zoek is? Registers zijn opgezet voor het vastleggen van administratieve concepten. Uit de administratieve concepten worden, door harmonisatie, statistische concepten afgeleid. Wanneer een administratief concept afwijkt van het statistische concept kan daar in de regel weinig aan veranderd worden (Bakker, 2009). Het is belangrijk dat men zich bewust is van de gevolgen en beperkingen van deze aanpak (Bakker, 2010).

Ook de dekking van de populatie is een belangrijk onderdeel: Komt de populatie van de bron overeen met de populatie die de onderzoeker wil bestuderen? Het standaardvoorbeeld hiervoor is natuurlijk dat mensen die illegaal in Nederland verblijven niet in de GBA zijn opgenomen. Ze wonen echter wel in Nederland en daar wil een onderzoeker graag ook iets over kunnen zeggen. Vaak zijn die niet-waargenomen of marginale groepen juist ook beleidsmatig interessant; dus niet alleen statistisch-beschrijvend (wie zijn het?) maar ook: wat doen ze? en hoe gaat het met ze? Registers zijn hier als zodanig uiteraard niet bruikbaar voor. Maar ook steekproeven uit een kader zijn dat niet, want dat kader is er nu juist niet. Hier zijn andere methoden voor nodig, zoals de vangst/hervangst methode (Sikkel et al., 2006). Deze wordt momenteel in een samenwerking tussen de Universiteit van Utrecht en het CBS nader onderzocht.

### 2.2.2. *Gaafmaken*

Het tweede onderzoeksveld betreft het gaafmaken, dat, zoals gezegd, bestaat uit het detecteren en corrigeren van verdachte waarden in registers. Sommige

controleregels zijn eenvoudig: een 12-jarige heeft nog geen vijf kinderen. Andere zijn echter complexer of gebaseerd op inhoudelijke kennis. Daar zijn dan vaak multivariate analysemethoden voor nodig om die fouten op te sporen. Bij beide typen regels is, zowel wat betreft inhoudelijke kennis als analysemethoden, inbreng van sociale wetenschappers noodzakelijk. Beide type regels worden bij het controleren en corrigeren van verdachte registerwaarden gebruikt. Dit is een onderwerp waar nog meer onderzoek nodig is (Hoogland et al., 2010).

### 2.2.3. *Combineren van bronnen*

Dit onderzoeksveld betreft het combineren van verschillende en soortgelijke bronnen, dus het combineren van registers met registers en registers met enquêtes. Hiervoor kunnen verschillende methoden worden gebruikt en hierbij treden verschillende problemen op. Voorbeelden hiervan zijn: gemiste koppelingen, miskoppelingen en selectiviteit.

Bij het gebruik van meerdere bronnen is correct combineren belangrijk. Het beoogde doel hiervan is immers verrijking van de data en dus kwaliteitsverbetering. Voordeel van registers is dat de eenheden in deze bronnen tegenwoordig steeds vaker, maar nog lang niet altijd, unieke koppel-identificatoren bevatten zoals het Burgerservicenummer (BSN). Helaas is dit bij enquêtedata vaak niet het geval. Gemiste en foutieve koppelingen kunnen het doel van verrijking verstoren en selectiviteit in het resterende gekoppelde bestand tot gevolg hebben. Een voorbeeld hiervan vinden we terug in het artikel van Anton Kunst, die de Landelijke Medische Registratie (LMR) gebruikt om de veranderingen van de gezondheidssituatie in Vogelaarwijken te analyseren (Kunst, 2010). Doordat in de LMR de BSN-nummers ontbreken, moet met kunst en vliegwerk worden gekoppeld. Ondanks dat dit met de nodige zorg gebeurt worden daarbij toch veel koppelingen gemist.

Soms geven registers en enquêtes over hetzelfde thema verschillende antwoorden. Het artikel van René Bekkers (2010) is daar een voorbeeld van. Hij laat zien dat er verschillen zijn tussen giften zoals ze in een enquête opgegeven worden en zoals ze in een register geregistreerd worden; gemiddeld komen ze overigens weer tot hetzelfde resultaat. De vraag of dit aan de enquête of aan het register ligt is overigens nog niet definitief beantwoord.

Het moge het duidelijk zijn dat we nog steeds enquêtes nodig hebben om de sociaalwetenschappelijke onderzoeksvragen te beantwoorden. Vaak is de registerinformatie nog te beperkt om met name verklarende vragen te kunnen beantwoorden. Zachtere variabelen zoals doorzettingsvermogen, houdingen en attitudes die een rol spelen in de verklaring van verbanden worden domweg niet geregistreerd. Een ander belangrijk voorbeeld is het opleidingsniveau, waarover Aslan Zorlu (2010) sprak in zijn bijdrage. Als we voor het bepalen van het opleidingsniveau alleen registers zouden gebruiken, komen we echt tekort; zeker de opleidingsgegevens van recente datum. Ook daarom moeten we gebruik blijven maken van enquêtes, in dit geval als aanvulling op de gegevens in beschikbare registers.

#### 2.2.4 *Schatten*

Het onderzoeksveld schatten met registers omvat vele onderzoeksvragen, zoals het consistent schatten op basis van meerdere bronnen, het bepalen van de (on)nauwkeurigheid van gecombineerde schatters, schattingen gebaseerd op zich geleidelijk vullende bronnen en het omgaan met methodebreuken.

#### *Combinatie van bronnen*

Er zijn verschillende manieren waarop er geschat kan worden bij combineren van steekproeven en registers, of bij het bijschatten voor onderdekking en gemiste koppelingen voor registers. Een veel gebruikte methode is het schatten met steekproefgewichten. Een andere methode is het modelmatig schatten van de parameters waarin men geïnteresseerd is. Een belangrijk voorbeeld van dit type schatters zijn kleinedomeinschatters. Kleine domeinen zijn deelpopulaties waarbij de steekproefomvang te klein is om betrouwbare directe (designgebaseerde) schattingen te maken. Het model, dat de verschillende domeinen met elkaar verbindt, eventueel door gebruik te maken van relevante hulpinformatie op domeinniveau, zorgt voor betere schattingen.

Allereerst speelt de vraag hoe consistente schattingen verkregen kunnen worden op basis van meerdere bronnen. Bijvoorbeeld de totale loonsom en de verdeling ervan over deelgroepen in het ene bestand moet gelijk zijn die in een ander bestand. Hierbij gelden soms ook nog aanvullende restricties. De centrale onderzoeksvraag is hierbij welke methode (herhaald wegen, herhaald imputeren of zelfs andere methoden zoals integratietechnieken of “gewone” imputatie) hiervoor gebruikt zou moeten worden. Overigens komen restricties vooral voor bij economische statistieken.

#### *Nauwkeurigheid van gecombineerde schatters*

De volgende vraag die speelt is: op welke wijze kan de onnauwkeurigheid van gecombineerde schatters worden bepaald? Voor de combinatie van registers en enquêtes zou dit gedaan kunnen worden door gebruik te maken van resamplingsmethoden zoals de bootstrap (Kuijvenhoven en Scholtus, 2010).

#### *Voorlopige schatters*

Sommige registers zijn nog niet compleet op het moment dat de data verwerkt moeten worden om op tijd bepaalde output te kunnen maken. Op een later tijdstip (bij latere releases) is die vulling wel adequaat of zelfs volledig, maar dat is pas na het gewenste publicatiemoment. Bronnen die het CBS gebruikt waarbij dit probleem zich concreet voordoet zijn een aantal administraties van de Belastingdienst. Dergelijke bronnen worden geleidelijk gevuld met cijfers al naargelang de aanslagen opgelegd worden. Dit fenomeen wordt ‘administratieve vertraging’ genoemd (Bakker, 2009) en is vooral voor het CBS belangrijk vanwege strikt vastgelegde en gehanteerde publicatiemomenten. Hier speelt een belangrijke kwaliteitseis van de

output een grote rol. Het CBS wil de verschillen in de schattingen (tussen releases) binnen bepaalde marges houden. Het gaat immers om de betrouwbaarheid/geloofwaardigheid van de cijfers. Bij een nieuwe schatting zullen de schattingen die op een eerdere release zijn gebaseerd moeten worden bijgesteld. Te hopen is dat deze bijstellingen klein zijn. Er zal dan een methode (model) moeten worden gekozen die rekening houdt met een eventuele selectiviteit van de vroege respondenten. Hiervoor kunnen bijvoorbeeld longitudinale modellen of tijdreeksmodellen worden gebruikt die rekening houden met selectieve informatie die veroorzaakt wordt door vroege en late respondenten. Een andere mogelijkheid is het gebruik van representativiteitsmaten. Hiermee kan men kwantificeren in welke mate vroege respondenten afwijken van de totale populatie. Dit kan vervolgens gebruikt worden om voor de selectiviteit van de vroege respondenten te corrigeren. Bovendien zouden Bayesiaanse schatters gebruikt kunnen worden om eerder gepubliceerde cijfers ('voorlopig definitief') te verrijken met gegevens die op een later tijdstip zijn binnengekomen zodat een beter definitief cijfer gepubliceerd kan worden. Wanneer de beoogde output microdata betreft, zal gewacht moeten worden tot het merendeel van de gegevens ontvangen is. De schattingen op geaggregeerd niveau kunnen (achteraf) voor het bepalen van dat moment gebruikt worden.

#### *Methodebreuken*

Een laatste deelterrein is dat van de methodebreuken. Veel onderzoeken worden herhaaldelijk in de tijd uitgevoerd. Hierdoor is het mogelijk om tijdreeksen op te bouwen die de ontwikkeling van een bepaald maatschappelijk verschijnsel beschrijven. Om er voor te zorgen dat deze reeksen consistent zijn, blijft de opzet van het onderzoek bij voorkeur zo lang mogelijk ongewijzigd. Het is echter onvermijdelijk om van tijd tot tijd de opzet van een onderzoek te herzien, bijvoorbeeld om de kwaliteit of de efficiëntie van het proces te verbeteren. Voorbeelden van dergelijke veranderingen zijn het overgaan van enquête- naar registergegevens of wijzigingen in de samenstelling van een register door de bronhouder. Dit heeft doorgaans tot gevolg dat er systematische effecten optreden in de cijfers die worden berekend. Om er voor te zorgen dat de continuïteit van tijdreeksen zo min mogelijk verstoord wordt, is het van belang dat het systematische effect van een dergelijke structurele wijziging zo goed mogelijk wordt gescheiden van de werkelijke ontwikkeling van de indicatoren. Deze doen zich bijvoorbeeld voor als de registerhouder de samenstelling of de vulling van het register wijzigt. Vergelijking met gegevens van een jaar eerder, is dan niet meer mogelijk. Dit bemoeilijkt longitudinale analyse of maakt het soms zelf onmogelijk. Dat dit een wezenlijk probleem is heeft André Wierdsma in zijn artikel laten zien. Voor vraagstukken in de geestelijke gezondheidszorg zijn goede longitudinale data nodig (Wierdsma, 2010). Ook Didier Fouarge (2010) heeft het belang van longitudinale data, voor het beschrijven van loopbanen op de arbeidsmarkt en in de sociale zekerheid, laten zien. Een oplossing voor methodebreuken is het scheiden van de methodewijziging en

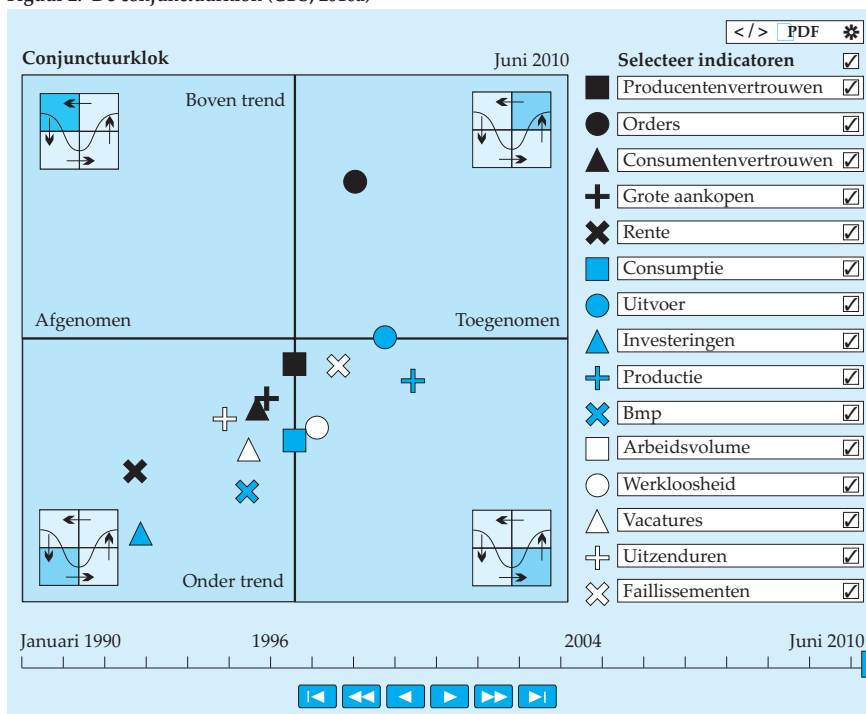
de werkelijke ontwikkeling. Dat kan lang niet altijd op macroniveau. Daarom zijn hiervoor door het CBS tijdreeksmodellen voor microdatabestanden ontwikkeld (Van den Brakel en Roels, 2009).

### 2.2.5 *Datavisualisatie*

Het laatste onderzoeksveld dat we bespreken is dat van de datavisualisatie. De uitdrukking “een plaatje zegt meer dan duizend woorden” illustreert dit. We kunnen immers in één oogopslag een gegevensvisualisatie overzien, terwijl het lezen van een tekst over hetzelfde onderwerp vaak veel meer tijd kost.

Voorbeelden van visualisatiemethoden zijn staafdiagrammen, puntenwolken en ruimtelijke visualisaties (kaarten). Relatief nieuw zijn bewegende visualisaties. Voorbeeld van de laatste is de conjunctuurklok van het CBS (Figuur 2). Hierin wordt de conjunctuurbeweging getoond aan de hand van de ontwikkeling (in de tijd) van een set van belangrijke conjunctuurindicatoren (CBS, 2010a). Een ander voorbeeld van een bewegende visualisatie is dat van de resterende (gezonde) levensverwachting (Figuur 3). In deze animatie varieert de bevolkingspiramide interactief al naargelang de gezichtspunten die worden gekozen (CBS, 2010b). Visualisatie is belangrijk voor output en analyse (Uwin et al., 2006), maar kan ook bij het gaafmaken worden gebruikt, bijvoorbeeld om uitbijters op te sporen in de puntenwolk (Hacking, 2009). Vandaag hebben we ook een voorbeeld gezien waarbij visualisatie gebruikt wordt voor de input (Daas et al., 2010). Echter, besef wel dat Datavisualisatie = Model: achter elke visualisatie zit een analysegedachte of -model. Er zal dus steeds zorgvuldig gekeken moet worden of de beelden die worden getoond, maar natuurlijk ook weer opgeroepen worden, methodologisch te verantwoorden zijn.

Figuur 2. De conjunctuurklok (CBS, 2010a)

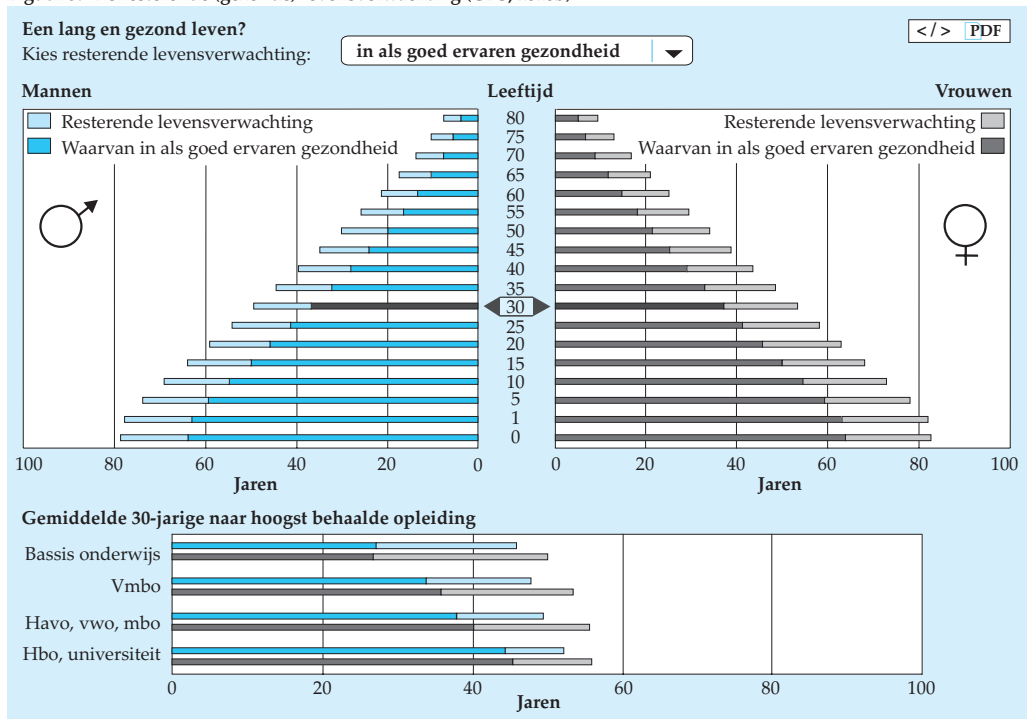


Bron: CBS

### 3. Samenwerking tussen het CBS en de universiteiten

We hebben gezien dat veel van de genoemde problemen en methoden gemeenschappelijk zijn tussen universiteiten en CBS. We dienen dan ook samen te zoeken naar oplossingen. Mogelijke samenwerkingsverbanden zijn te vinden op het niveau van data-infrastructuur, gezamenlijk onderzoek en beleid. Het CBS kan onderzoekers van universiteiten op een aantal manieren ondersteunen bij het toegankelijk maken van (beveiligde) gegevens uit registers. De eerste is via Data Archiving and Networked Services (DANS), een initiatief van de Nederlandse Organisatie van Wetenschappelijk Onderzoek en de Koninklijke Nederlandse Academie van Wetenschappen. Tussen het CBS en DANS bestaat een overeenkomst over de beschikbaarstelling van bestanden met gegevens over personen en huishoudens voor wetenschappelijk onderzoek. Hierdoor kan DANS zogenaamde 'beveiligde microbestanden,' bestanden die zodanig bewerkt zijn dat herkenning van individuen en huishoudens uitgesloten is, onder een aantal voorwaarden aan universiteiten en onderzoeksinstituten beschikbaar stellen.

Figuur 3. De resterende (gezonde) levensverwachting (CBS, 2010b)



Bron: CBS

De tweede mogelijkheid is via het zogenaamde ‘on-site’ werken. Individuele onderzoekers, van een aantal gemachtigde organisaties, kunnen, na toestemming, zelf aan de slag met CBS-gegevens in speciaal daarvoor ingerichte werkruimtes in Den Haag en Heerlen. Voor gemachtigde onderzoekers kan dit tegenwoordig zelfs via remote access. Op de website van het Centrum voor Beleidsstatistiek van het CBS (2010c) is hier meer informatie over te vinden. Samenwerking tussen universiteiten en het CBS vindt tevens plaats door het aanstellen van hoogleraren, promovendi en stagiaires. Hierdoor vindt steeds meer gezamenlijk onderzoek plaats. In deze bundel vindt men hier al een aantal voorbeelden van. De methodologische onderzoeksvragen die naar aanleiding van de bespreking van het statistische proces naar voren komen leveren bijzonder interessante onderzoeksonderwerpen, ook voor sociale wetenschappers. Onderzoek naar het gebruik van registerdata is een gezamenlijk element voor statistiek en wetenschap en hier kunnen en moeten we samen optrekken.

## Referenties

- Bakker, B. (2009) *Trek alle registers open!* Inaugurale rede, Faculteit der Sociale Wetenschappen, Vrije Universiteit Amsterdam.
- Bakker, B. (2010) Vaststellen van de validiteit van registervariabelen. *Registers in sociaalwetenschappelijk onderzoek: Mogelijkheden en valkuilen*, pp. 15–30.
- Bekkers, R. (2010) Nauwkeurigheidsmetingen in registers en enquêtes. *Registers in sociaalwetenschappelijk onderzoek: Mogelijkheden en valkuilen*, Leiden, pp. 75–88.
- Bethlehem, J. (2008) Surveys without questions. In *International Handbook of Survey Methodology*, eds. E.D. de Leeuw, J.J. Hox, D.A. Dillman, New York: Lawrence Erlbaum Associates, pp. 500–511.
- Bijleveld, C. (2010) Registers en criminologisch onderzoek. *Registers in sociaalwetenschappelijk onderzoek: Mogelijkheden en valkuilen*, pp. 65–74.
- CBS (2010a) Webpagina conjunctuurklok. CBS-website, <http://www.cbs.nl/nl-NL/menu/themas/dossiers/conjunctuur/publicaties/conjunctuurbericht/inhoud/conjunctuurklok/conjunctuurklok2.htm>.
- CBS (2010b) Webpagina resterende (gezonde) levensverwachting. CBS-website, <http://www.cbs.nl/nl-NL/menu/themas/gezondheid-welzijn/cijfers/extra/resterende-gezonde-levensverwachting.htm>.
- CBS (2010c) Nieuwspagina Centrum voor Beleidsstatistiek. CBS-website, <http://www.cbs.nl/nl-NL/menu/informatie/beleid/nieuws/default/htm>.
- Daas, P.J.H., Ossen, S.J.L., Tennekes, M. (2010) *The determination of administrative data quality: recent results and new developments*. Paper for the European Conference on Quality in Official Statistics 2010, Helsinki, Finland.
- De Wolf, P-P., van Delden, A. (2009) *Onderzoeksvragen registers voor de komende jaren*. Interne CBS-nota, DMV-2009-01-05-PWOF-ADLN.
- Eurostat (2009) ESS standard for quality report. Eurostat Methodologies and Working papers, Office for Official Publications of the European Communities, Luxembourg.
- Fouarge, D., van Gaalen, R., de Grip, A. (2010) Participatie en loon na bedrijfs-economisch ontslag kapitaalvernietiging of negatief signaal? *Registers in sociaalwetenschappelijk onderzoek: Mogelijkheden en valkuilen*, pp. 31–50.



Gemeente Amsterdam (2008) *Kwaliteit van de Amsterdamse GBA is vooruitgegaan, maar moet beter*. Persbericht Dienst Persoonsgegevens, Gemeente Amsterdam, 13 juni.

Hacking, W. (2009) *Macro-selection and micro-editing: a prototype*. Paper prepared for the 12<sup>th</sup> International Blaise Users Conference (IBUC), Riga, Latvia, pp. 118–125.

Hoogland, J., van der Loo, M., Pannekoek, J., Scholtus, S. (2010) *Controle en correctie*. CBS-methodenreeks, Statistische methoden 10011, Den Haag: Centraal Bureau voor de Statistiek.

Hox, J.J., Boeije, H.R. (2005) Data Collection, Primary vs. Secondary. *Encyclopedia of Social Measurement*, Vol. 1, ed. K. Kempf-Leonard, London: Academic Press, pp. 593–599.

Kuijvenhoven, L, Scholtus, S. (2010) *Estimating accuracy for statistics based on register and survey data*. Discussion paper 10007, The Hague/Heerlen: Statistics Netherlands.

Kunst, A. (2010) Koppeling van registers: Onmisbaar voor een beter overzicht in de volksgezondheid. *Registers in sociaalwetenschappelijk onderzoek: Mogelijkheden en valkuilen*, pp. 93–110.

Sikkel, D., van der Heijden, P.G.M., van Gils, G. (2006). *Methoden voor omvangschattingen van verborgen populaties, met name illegalen*. WODC rapport 243, Onderzoek en beleid. Meppel: Boom Juridische Uitgevers.

Uwin, A., Theus, M., Hofmann, H. (2006) *Graphics of Large Datasets: Visualizing a Million*. Singapore: Springer.

Van den Brakel, J., Roels, J. (2010) *Intervention analysis with state space models to estimate discontinuities due to a survey redesign*. Discussion paper 09039, The Hague/Heerlen: Statistics Netherlands.

Wallgren, A., Wallgren, B. (2007) *Register-based Statistics: Administrative Data for Statistical Purposes*. Chichester: John Wiley & Sons.

Wierdsma, A. (2010) Continuïteit van zorg: Indicatoren voor lange termijn zorg en zekerheid in de GGZ. *Registers in sociaalwetenschappelijk onderzoek: Mogelijkheden en valkuilen*, pp. 111–126.

Willeboordse, A. (2008) *Inleiding in het Statistisch proces en de CBS-methodenreeks*, Voorburg: Centraal Bureau voor de Statistiek.

Zorlu, A. (2010) Opleidingsniveau in registers: Een toets van de validiteit via loonfuncties. *Registers in sociaalwetenschappelijk onderzoek: Mogelijkheden en valkuilen*, pp. 51–64.