REGIOPLAN

METHODENSTUDIE METING NALEVINGSNIVEAUS SOCIALE ZEKERHEID

- eindrapport -

Dr. S. van der Ploeg
Drs. J. Soethout
Drs. B. van Waveren

Amsterdam, april 2006
Regioplan publicatienr. 1396

Regioplan Beleidsonderzoek
Nieuwezijds Voorburgwal 35
1012 RD Amsterdam
Tel.: +31 (0)20 - 5315315
Fax : +31 (0)20 - 6265199

Onderzoek uitgevoerd door Regioplan Beleidsonderzoek in opdracht van het Ministerie van Sociale Zaken en Werkgelegenheid.

INHOUDSOPGAVE

Voorwoord

1 Inleiding ... 1
- 1.1 Achtergrond ... 1
- 1.2 De verplichtingen ... 1
- 1.3 Doel en vraagstelling onderzoek ... 4
- 1.4 Onderzoeksaanpak ... 5
- 1.5 Beoordelingscriteria ... 6
- 1.6 Opbouw van het rapport ... 7

2 Verkenning onderzoek zelfrapportage ... 9
- 2.1 Inleiding ... 9
- 2.2 Beschrijving belangrijke kenmerken huidige methodiek ... 9
- 2.3 Beoordeling huidige methodiek ... 10
- 2.4 Samenvattend overzicht ... 23

3 Aselecte controles en DCE ... 27
- 3.1 Inleiding ... 27
- 3.2 Controles ... 27
- 3.3 Ervaringen in de praktijk ... 29
- 3.4 Bruikbaarheid aselecte controles ... 31
- 3.5 Detection Controlled Estimation ... 34
- 3.6 Samenvattend overzicht ... 36

4 Administraties ... 41
- 4.1 Inleiding ... 41
- 4.2 De sollicitatieverplichting (WWB, WW) ... 41
- 4.3 Trajectverplichting (WWB, WW, WGA) ... 43
- 4.4 Opgave inkomsten (WWB, WW, WIA, AOW Anw ... 44
- 4.5 Opgave gezamenlijke huishouding (WWB, AOW, Anw) ... 47
- 4.6 Melden herstel (WIA) ... 48
- 4.7 De methodiek beoordeeld ... 49

5 Methoden en verplichtingen ... 51
- 5.1 Inleiding ... 51
- 5.2 Verplichting tot solliciteren ... 51
- 5.3 Nakomen trajectafspraken ... 55
- 5.4 Opgave van inkomsten uit arbeid ... 58
- 5.5 Verplichting opgeven gezamenlijke huishouding ... 62
- 5.6 Melden herstel, meewerken aan herstelactiviteiten ... 65

6 Afsluiting ... 69
- 6.1 Inleiding ... 69
- 6.2 Overzicht van prestaties van methoden naar verplichting ... 69

6.3 Vergelijking van de methodes: bespreking 70
6.4 Conclusie 72
6.5 Uitgangspunten bij opstellen advies 72
6.6 Advies POROSZ 2006 en verder 72

Literatuur 77

Bijlagen…. 79
Bijlage 1 Informanten 81
Bijlage 2 Tabel nalevingsniveaus 83

VOORWOORD

Regelovertreding maakt regelingen nodeloos duur en knaagt aan het draagvlak voor de instandhouding ervan. Het bevorderen van de naleving van de regels is voor het Ministerie van SZW een belangrijke zaak, zelfs in die mate, dat het als prestatie-indicator in de begroting van het ministerie is opgenomen. Het vaststellen van de nalevingsniveaus gebeurt met behulp van de zogenaamde *randomized-responsemethode*. Deze methode wordt gekenmerkt door een sterke nadruk op de waarborg voor anonimiteit van de respondent. De gedachte daarachter is dat deze waarborgen de bereidheid van de respondent bevorderen om laakbaar gedrag toe te geven. Maar deze techniek heeft ook nadelen, waaronder het kostenaspect en de ruime betrouwbaarheidsmarges.

Tegen die achtergrond is binnen het Ministerie van SZW de behoefte ontstaan om de gebruikswaarde van de huidige onderzoeksmethode nader te bepalen en na te gaan of er aantrekkelijke alternatieven voorhanden zijn.
Het Ministerie heeft Regioplan Beleidsonderzoek de opdracht verleend om een dergelijk onderzoek uit te voeren. Voor u ligt het verslag van deze studie.

Voor dit onderzoek is gesproken met een keur aan deskundigen op het terrein van methodologie en van regelovertreding. We zijn hen zeer dankbaar voor hun medewerking.

Voor de begeleiding van het onderzoek is door de opdrachtgever een klankbordgroep ingesteld, bestaande uit:

- de heer P. Oomens, Ministerie van SZW (voorzitter);
- de heer K. de Wit, Ministerie van SZW;
- de heer A. de Graaf, Ministerie van SZW;
- de heer M. Einerhand, Ministerie van SZW;
- de heer G. Duinkerken, Monitorgroep/Steunpunt Ketensamenwerking;
- de heer H. Aukes, UWV;
- de heer F. Werner, SVB;
- de heer W. Schaasberg, CBS;
- de heer V. Neevel, Ministerie van VROM;
- de heer P. Lieffering, Ministerie van VROM.

We danken alle leden voor de constructieve en prettige manier waarop de begeleiding ter hand is genomen.

Amsterdam, 21 april 2006,

Bob van Waveren
projectleider

1 INLEIDING

1.1 Achtergrond

Regelovertreding maakt regelingen nodeloos duur en knaagt aan het draagvlak voor de instandhouding ervan. Dat geldt ook in de sociale zekerheid. Het bevorderen van de naleving van de regels is voor het Ministerie van Sociale Zaken en Werkgelegenheid (SZW) een belangrijke zaak, zelfs in die mate dat het als prestatie-indicator in de begroting van het ministerie is opgenomen. Het vaststellen van de nalevingsniveaus gebeurt met behulp van de zogenaamde *randomized-responsemethode*. Deze methode wordt gekenmerkt door een sterke nadruk op de waarborg voor anonimiteit van de respondent. De gedachte daarachter is dat deze waarborgen de bereidheid van de respondent bevorderen om laakbaar gedrag toe te geven.
Maar deze techniek heeft ook nadelen. In de startnotitie noemt het Ministerie van SZW het kostenaspect, de ruime betrouwbaarheidsmarges en het aspect van onderrapportage. Dat heeft onder meer tot gevolg dat het lastig is om ontwikkeling in de loop der tijd te kunnen interpreteren, aangezien veel van de verschillen binnen de marges blijven en derhalve niet statistisch significant zijn.
Tegen die achtergrond is binnen het Ministerie van SZW de behoefte ontstaan om de gebruikswaarde van de huidige onderzoeksmethode nader te bepalen en na te gaan of er aantrekkelijke alternatieven voorhanden zijn. Daarbij gaan de eerste gedachten uit naar het gebruik van registraties op terrein van controle, opsporing en fraude. In aanvulling daarop heeft men ook interesse in mogelijkheden in de vorm van alternatieve vormen van persoonsonderzoek.

Regioplan Beleidsonderzoek heeft de gebruikswaarde van de randomized-responsemethode en van andere onderzoeksmethoden onder de loep genomen. Het verslag van dit onderzoek ligt voor u.
In dit eerste hoofdstuk beschrijven we van welke verplichtingen de nalevingsniveaus opgenomen dienen te worden als prestatie-indicator in de ministeriële begroting. Ook geven we aan welke vraagstelling bij het onderzoek geldt en welke onderzoeksaanpak wij gehanteerd hebben bij het beantwoorden van de vragen. We besluiten dit hoofdstuk met een beschrijving van de opbouw van het rapport.

1.2 De verplichtingen

Prestatie-indicatoren naleving
In de jaarlijkse begroting presenteert het Ministerie van SZW haar voornemens op het terrein van de sociale zekerheid. Deze worden onder meer zichtbaar gemaakt met de zogenaamde presentatie-indicatoren. Dat gebeurt ook waar

het gaat om de naleving van de wet- en regelgeving op het terrein van de Bijstandswet, WW, WAO/WIA, AOW, AKW en Anw. Voor elk van deze wetten apart wordt het gerealiseerde nalevingsniveau weergegeven over het voorgaande jaar en wordt een norm gesteld voor het begrotingsjaar.

Aard van de verplichtingen
De naleving heeft betrekking op een aantal, al dan niet wetspecifieke verplichtingen. We vatten deze samen in onderstaande tabel. Enkele van deze indicatoren zijn niet in de SZW-begroting 2006 van opgenomen.[1]

Tabel 1.1 Relevante verplichtingen voor uitkeringsontvangers sociale zekerheid

	WWB	WW	WAO/WIA*		AOW	Anw	AKW
			IVA	WGA			
Verplichting tot:							
Opgave van inkomsten	X	X	X	X	X	X	
Opgave inkomsten kind >16 jr							X
Solliciteren	X	X					
Trajectverplichtingen	X	X		X			
Verplichtingen i.h.k.v. herstel			X	X			
Opgeven gezam .huishouding	X				X	X	

* De WAO is vervangen door de WIA. De WIA valt uiteen in twee regelingen: de IVA voor volledig arbeidsongeschikte werknemers en de WGA voor gedeeltelijk arbeidsongeschikte werknemers.

We geven een korte toelichting op iedere verplichting die in het onderzoek betrokken is.

Opgave van inkomsten
Voor alle genoemde regelingen geldt dat het recht op en de hoogte van de uitkering afhankelijk is van het inkomen dat de uitkeringsontvanger uit andere bronnen heeft. Inkomsten dienen dan ook direct te worden gemeld. Recht op en hoogte van de uitkeringen WWB, AOW en Anw is afhankelijk van het gezinsinkomen en daarmee van het inkomen van een eventuele partner. Voor de AKW zit het net iets anders in elkaar, omdat hier niet het inkomen van de uitkeringsontvanger of de partner telt, maar het inkomen van zijn of haar kind, in geval dit ouder is dan zestien jaar en thuis woont.

De sollicitatieverplichting
De werkloosheidsregelingen (WWB en WW) ondersteunen mensen in geval van werkloosheid. Op ontvangers van werkloosheidsuitkeringen rust de verplichting om zo snel als mogelijk weer aan het werk te komen. De

[1] Niet opgenomen in de begroting 2006: naleving van de sollicitatieverplichting WWB; opgave van inkomsten Anw en AOW, naleving van reïntegratieverplichtingen.

sollicitatieverplichting is daarvan een prominent voorbeeld. Op basis van individuele omstandigheden kunnen cliënten (tijdelijk) ontheven worden van de sollicitatieverplichting.

Verplichting tot het volgen van een re-integratietraject
Ten behoeve van het onderzoek operationaliseren we het begrip re-integratie-traject als een vanuit de uitkeringsorganisatie (UWV, gemeente) aangeboden voorziening, gericht op de arbeidsinschakeling (of sociale activering). In een trajectplan wordt overeengekomen en vastgelegd waaruit deze voorziening bestaat. WWB- en WW-gerechtigden zijn verplicht om mee te werken aan de uitvoering van dit trajectplan.
Niet alle uitkeringsontvangers krijgen zo´n re-integratietraject aangeboden. Ten eerste is er een groep van wie wordt verwacht dat zij op eigen kracht weer aan de slag komt. Ten tweede kunnen er gronden zijn om uitkeringsgerechtigden vrijstelling van een of meer van de arbeidsverplichtingen te verlenen, bijvoorbeeld vanwege persoonlijke belemmeringen of zorgtaken. Ten derde kan het zijn dat het traject in voorbereiding is: de cliënt zit op op dat moment nog niet op een traject, maar enige tijd later vermoedelijk wel.

Melding in het kader van herstel
De WAO bood ondersteuning in geval van verlies van verdiencapaciteit die voortkomt uit al dan niet gedeeltelijke arbeidsongeschiktheid. Per januari 2006 is de WAO opgevolgd door de WIA. Herstel kan gevolgen hebben voor het recht op of de hoogte van de uitkering en dient dus direct te worden gemeld. Ook zijn WAO'ers verplicht mee te werken aan alles dat kan bijdragen tot het herstel. Dus moeten ze meewerken aan medische behandelingen en/of andere op herstel gerichte handelingen. Meldingen kunnen ook hierop betrek-king hebben.

Melding van gezamenlijke huishouding
Hoogte van de uitkeringen WWB, AOW en Anw wordt gedifferentieerd naar huishoudsamenstelling. Omdat het alleenstaanden en eenoudergezinnen ontbreekt aan zekere schaalvoordelen kan hen een toeslag worden verleend. Gevolg daarvan is dat twee alleenstaanden per saldo een hogere uitkering ontvangen dan een (echt)paar. Dit kan mensen in de verleiding brengen om hun huishoudsituatie (op papier) anders voor te doen dan in werkelijkheid.
In de praktijk is het lastig om vast te stellen of er daadwerkelijk sprake is van een gezamenlijke huishouding. Het criterium van de economische verstren-geling dat hiervoor geldt, is voor uitkeringsontvangers moeilijk te vatten en voor uitvoeringsinstanties lastig te bewijzen.

1.3 Doel en vraagstelling onderzoek

Doel

Doel van het onderzoek is het verkrijgen van een onderbouwde beschrijving van de meest adequate aanpak voor het bepalen van de nalevingsniveaus van de onderscheiden wetten. De meest adequate aanpak kiezen we uit een serie van opties, waaronder de administratieve aanpak, aselecte controles, de randomized-responsemethode (inclusief mogelijke verbeteringen) en eventuele alternatieve vormen van persoonsonderzoek.

Beperkingen

Het testen van de voorgestelde aanpak valt buiten het kader van dit onderzoek. Het gaat in dit project dus om de ontwikkeling van een alternatief onderzoeksmodel en niet om de uitvoering van een onderzoek. Het onderzoek beperkt zich tot de in Nederland woonachtige uitkeringsontvangers.[2]

Probleemstelling

We leiden hieruit de volgende *probleemstelling* af:
Welke alternatieven zijn er denkbaar voor het gebruik van de randomized-responsemethodiek ter bepaling van de nalevingsniveaus? Hoe moet de gebruikswaarde van de huidige POROSZ-resultaten vanuit deze eventuele alternatieven worden beoordeeld?

Onderzoeksvragen

Wij formuleren de volgende *onderzoeksvragen*.

1. *Optimalisering randomized response. Wat zijn de mogelijkheden om de randomized-responsemethodiek te verbeteren, met name op het punt van de referentiegroepen?*
   Hoe kan dat worden uitgewerkt en wat zijn de eventuele neveneffecten?
   In de huidige berekening van de prestatie-indicatoren voor naleving van de werknemersverzekeringen en de WWB wordt uitgegaan van de gehele groep uitkeringsgerechtigden en bij de volksverzekeringen van een beperkte selectie. Het is de vraag hoe dit doorwerkt op de resultaten van het Periodiek Onderzoek Regelovertreding Sociale Zekerheid (POROSZ), hoe dit moet worden beoordeeld en welke voorstellen tot aanpassingen daaruit kunnen worden afgeleid. Denkbaar is bijvoorbeeld, om bij de selectie meer precies uit te gaan van de groepen mensen die reëel de mogelijkheid hebben een bepaalde regel te overtreden. Dat roept wel weer vragen op, bijvoorbeeld over de criteria die hierbij moeten worden gehanteerd. Hoewel met name het punt van de referentiegroepen in de onderzoeksvraag wordt benoemd, bestuderen we bij verbetermogelijkheden van de methodiek alle mogelijke zaken die daarbij relevant zijn en aan het licht komen.

2. *Alternatieve methoden voor zelfrapportage. Zijn er andere methoden bij zelfrapportage dan de randomized-responsemethode, om op een verantwoorde wijze tot uitspraken te komen over de omvang van frauderend gedrag?*

[2] Een aanzienlijk aantal klanten van de SVB woont in het buitenland (ongeveer 250.000), bij UWV zijn het er ongeveer 30.000.

Op zijn minst is denkbaar, dat er andere methoden bestaan voor het bevragen van de doelgroep zelf (zelfrapportage), die een alternatief kunnen vormen voor de randomized-responsemethodiek. In het onderzoek achterhalen we deze alternatieven en verkennen we hun gebruikswaarde.

3. *Vervanging door controles. Zijn er met behulp van (aselecte) controles kwantitatieve uitspraken te doen die het mogelijk maken de ontwikkeling wat betreft de naleving van de wet te volgen?*
   Deze vraag gaat over de mogelijkheden om op basis van controles bij een deel van de populatie uitspraken te doen over de mate van naleving door de gehele populatie. Onder welke voorwaarden is dit mogelijk? Is het mogelijk bij alle relevante nalevingsniveaus?

4. *Gebruik administratieve data. Zijn er met behulp van administratieve gegevens kwantitatieve uitspraken te doen die het mogelijk maken de ontwikkeling wat betreft de naleving van de wet te volgen? Is het mogelijk om met behulp van administratieve data de uitkomsten van randomized-responseonderzoek te valideren?*
   De volgende vraag behandelt de mogelijkheden om de databronnen in te zetten naast randomized-responseonderzoek ter validering, of als zelfstandig hanteerbaar vervangend alternatief voor de randomized-responsemethodiek. Daarvan kan pas sprake zijn als blijkt dat dit alternatief in staat is om stabiele en betrouwbare cijferreeksen te leveren.

5. *Hoe moet de gebruikswaarde van de huidige POROSZ-resultaten vanuit de mogelijke alternatieven worden beoordeeld? Wat kan gelden als de meest adequate aanpak voor het meten van de nalevingsniveaus?*
   Ten slotte zet de laatste vraag aan tot een beoordeling van de beschikbare mogelijkheden en een keuze van de meest adequate aanpak. Indien nodig en nuttig wordt hierbij gedifferentieerd naar de verschillende wetten en/of items.

1.4 Onderzoeksaanpak

Om tot beantwoording van de onderzoeksvragen te komen, hebben wij verschillende onderzoeksmethoden gehanteerd, te weten:

- Literatuurstudie
- Documentanalyse
- Interviews

Literatuurstudie

Tijdens de literatuurstudie zijn verschillende publicaties over methoden van persoonsonderzoek en het schatten van ‘dark figures’ bestudeerd. Het betrof onder andere literatuur over randomized response, Detection Controlled Estimation (DCE), en over responsproblemen bij zelfrapportage.

In bijlage 1 is een overzicht van de geraadpleegde literatuur opgenomen.

Documentanalyse
Om meer zicht te krijgen op de ervaringen van uitkeringsinstanties met het gebruik van administraties en het uitvoeren van aselecte controles, is documentatie daarover bestudeerd. Het gaat dan bijvoorbeeld om documentatie van de uitkeringsinstanties zelf (statistiekbeschrijvingen, handhavingsplannen en -rapportages), maar ook om rapporten van de Inspectie voor Werk en Inkomen.

Interviews
De informatie uit het literatuuronderzoek en de documentanalyse is verdiept door middel van interviews met landelijke experts, zowel op het terrein van zelfrapportage en randomized response als op het terrein van handhaving in de sociale zekerheid. Wij hebben ook gesproken met medewerkers van de uitkeringsorganisaties, de Belastingdienst, het Inlichtingenbureau en het CBS over de beschikbaarheid en bruikbaarheid van administraties en over de ervaringen met aselecte controles. In bijlage 2 treft u een overzicht aan van de informanten die geïnterviewd zijn.

1.5 Beoordelingscriteria

Het Ministerie van SZW heeft vooraf geen strikte criteria voor een beoordeling van de huidige onderzoeksaanpak aangegeven. Regioplan heeft daarom uit de startnotitie met SZW en de gevoerde gesprekken zelf een aantal criteria gedestilleerd en daarop enkele aanvullingen gegeven.

Belangrijke criteria voor het beoordelen van onderzoek naar niet-naleving zijn wat ons betreft:
- Validiteit uitkomsten: de percentages overtreders die in het onderzoek worden gepresenteerd, moeten zo dicht mogelijk bij de werkelijke percentages liggen.
- Betrouwbaarheid uitkomsten: de percentages overtreders moeten zo precies mogelijk zijn (de betrouwbaarheidsmarges zo klein mogelijk).
- Bereik: het onderzoek zou op zo veel mogelijk vormen van niet-naleving toegepast moeten worden (toepasbaar kunnen zijn) en zo veel mogelijk informatie over het niet-naleving moeten opleveren.
- Doorlooptijd: de resultaten van een meting met als peiljaar t zouden in het jaar t+1 uiterlijk in februari beschikbaar moeten zijn voor opname in het jaarverslag over het jaar t.
- Privacybescherming: knelpunten ter zake van privacyaspecten.

Daarnaast zijn meegenomen de aspecten:
- Draagvlak in veld: de knelpunten rond het draagvlak.
- Kosten: indicatie van de kosten voor het onderzoek.

1.6 Opbouw van het rapport

In het vervolg van dit rapport beschrijven we eerst de bevindingen ten aanzien van randomized response en overige vormen van zelfrapportage in hoofdstuk 2. Mogelijkheden om met behulp van aselecte controles en Detection Controlled Estimation nalevingsniveau's in kaart te brengen, beschrijven wij in hoofdstuk 3. De administratieve mogelijkheden staan in hoofdstuk 4 centraal. In deze hoofdstukken worden de methoden steeds aan de hand van een vaste set van criteria beoordeeld. We besluiten het rapport met een hoofdstuk waarin de conclusies zijn weergeven. Ook wordt in het laatste hoofdstuk een advies gegeven ten aanzien van de toekomstige meting van nalevingsniveaus in de sociale zekerheid.

2 VERKENNING ONDERZOEK ZELFRAPPORTAGE

2.1 Inleiding

Binnen het totale project neemt de evaluatie van de huidige wijze van het meten van niet-naleving via het POROSZ een speciale plaats in. We geven in dit onderdeel een beschrijving van de essentie van de huidige methodiek en richten ons op de sterke en zwakke kanten ervan. Het is daarbij van belang te bedenken dat de vormgeving van het POROSZ op dit moment een uitkomst is van een proces van voorstellen, aanpassingen en afspraken die de uitvoerders (UU/BOA) en opdrachtgever (SZW) gezamenlijk hebben doorlopen. De uitvoerders wijzen er in dit verband op dat zwakkere elementen in de opzet door henzelf worden onderkend maar ook zijn doorgegeven en geaccepteerd door de opdrachtgever.

2.2 Beschrijving belangrijke kenmerken huidige methodiek

Sinds 2000 laat het Ministerie van SZW het Periodiek Onderzoek Regelovertreding Sociale Zekerheid (POROSZ) uitvoeren. Dit onderzoek heeft onder meer tot doel het niveau van niet-naleving vast te stellen. Het onderzoek wordt uitgevoerd door een samenwerkingsverband van BOA en de UU. Na enkele proefmetingen zijn de jaren 2000, 2002 en 2004 volwaardige POROSZ-metingen uitgevoerd.

De metingen 2000, 2002 en 2004 hebben betrekking op regelovertreding in de WAO, WW en Abw/WWB. In 2004 zijn daaraan de volksverzekeringen toegevoegd: AOW, AKW en Anw.

Het gaat om een persoonsonderzoek waarbij via zelfrapportage de mate van niet-naleving wordt bepaald. Omdat het om gevoelige vragen gaat, wordt een specifieke methodiek toegepast (randomized response) om respondenten een veilige omgeving te bieden voor eerlijke beantwoording (omdat niet iedereen bereid zal zijn fraude toe te geven in een vragenlijst). Dat wordt gedaan door respondenten (op de computer) met dobbelstenen te laten gooien en afhankelijk van die uitkomst een eerlijk of een gedwongen antwoord te laten geven. De afzonderlijke dobbelsteenuitkomst is voor de onderzoekers niet bekend. Individuele antwoorden hebben dus voor de onderzoekers geen betekenis. Dat is precies de anonimiteit en veiligheid waar de respondenten behoefte aan hebben. Op grond van de te verwachten dobbelsteenuitkomsten (die bekend zijn) kunnen op deze manier op geaggregeerd niveau wel goede schatters verkregen worden van de mate van naleving.

De randomized-responseaanpak is in Nederland (ook door het POROSZ-onderzoek) volop in ontwikkeling en daarmee loopt Nederland internationaal voorop in het ontwikkelen en toepassen van zelfrapportage voor fraudegedrag.

Bij de laatste twee metingen is gebruikgemaakt van zogenaamde access pools van TNS NIPO en GfK Panel Services. Gegevensverzameling gaat via elektronische vragenlijsten die naar respondenten verzonden worden en die ingevuld weer terug worden gestuurd (respondenten loggen dus niet in op een website). Deze panels zijn 'representatief op de binnen onderzoek gebruikelijke achtergrondkenmerken' volgens het POROSZ-rapport. TNS NIPO geeft aan dat de pools geschikt zijn voor het trekken van adequate steekproeven. Deelnemers ontvangen een vergoeding (spaarsysteem) voor deelname aan onderzoeken. Deelnemers aan de pools worden geworven via lopende telefonische ad-hoconderzoeken en er worden speciale wervings-acties (via brieven) gehouden voor het vullen van moeilijker bereikbare groepen.

Op grond van bekende kenmerken (en deels door selectievragen vooraf te stellen) konden uit deze access pools respondenten worden geselecteerd die tot de groep uitkeringsontvangers behoorden. In 2004 zijn voor WAO, WW en WWB per wet ruim 1500 respondenten uitgenodigd deel te nemen. Dat leverde een bruikbare respons van rond de 850 op. Voor de AOW zijn ruim 2500 en voor de AKW bijna 1300 respondenten uitgenodigd om deel te nemen met een bruikbare respons van 800. Voor de Anw is bijna 900 aan bruikbare respons gerealiseerd. Daarvoor zijn 2300 mensen uitgenodigd deel te nemen. Dit lijken relatief lage responspercentages voor gebruik van een access pool. De onderzoekers geven aan dat de selectievragen niet altijd aan de beoogde respondent maar vaak ook aan een ander lid van het huishouden is gesteld. Daardoor bleek een ongeveer kwart van de aangeschrevenen geen uitkering ontving of niet de uitkering ontving waarop door de onderzoekers was gerekend.

2.3 Beoordeling huidige methodiek

We behandelen in deze paragraaf de uitvoering van het huidige POROSZ-onderzoek aan de hand van de criteria die we in de inleiding hebben gepresenteerd om een beoordeling van de huidige metingen te kunnen geven.

2.3.1 Validiteit

In een enquêteonderzoek als het POROSZ kunnen we een aantal soorten fouten onderscheiden (bijv. Fox en Tracy, 1986). De volgende soorten fouten zijn hier met name relevant:

A Steekproeffouten
B Niet-steekproeffouten:
  B1 Non-responsafwijking
  B2 Reponsafwijking

Ad A: Steekproeffouten
Bij de steekproeffouten speelt de vraag of de mensen die in de steekproef zijn opgenomen een afspiegeling vormen van de populatie op de doelvariabele: zijn er redenen om aan te nemen dat de steekproef uit mensen bestaat die beter (of juist slechter) de regels naleven dan dat in de populatie het geval is? Door gebruik van de access pools is de voor het onderzoek beschikbare steekproef afhankelijk van de beslissingen van mensen die door de markt-onderzoekbureaus worden uitgenodigd deel te nemen aan het panel. Momenteel is de (impliciete) aanname in POROSZ dat niet-nalevers net zo positief tegenover deelname aan access pools van TNS NIPO en GfK staan als nalevers. We kunnen op grond van beschikbare informatie geen afdoend oordeel vellen over die aanname. Wel is voorstelbaar dat mensen die minder gevoelig zijn voor maatschappelijke druk en minder gezagsgetrouw zijn (en dus eerder geneigd zijn regels te overtreden) ook minder snel geneigd zijn mee te werken aan onderzoek (waarbij ook een beroep wordt gedaan op het maatschappelijk belang). Stoop (2005, p. 81) meldt op grond van internationaal onderzoek dat het meewerken aan enquêtes samenhangt met de mate waarin mensen sociaal zijn geïntegreerd. Daarnaast zijn er bevindingen dat mensen die meer uit zijn op persoonlijk gewin en minder op het maatschappelijk belang vaker weigeren mee te doen aan enquêtes (idem, p. 82). Dit zou tot een onderschatting van het niet-nalevingspercentage leiden wanneer enquêtes worden gebruikt om de mate van naleving te bepalen. We beschikken niet over gegevens die voldoende bewijs bieden om te concluderen dat op dit onderdeel de steekproef die wordt gebruikt in POROSZ systematisch zou afwijken. Het lijkt ons wel een taak van de onderzoekers (dan wel de panelbeheerders) om zorg op dit punt weg te nemen. Wat nu bekend is, is dat er buiten ophoging van de te benaderen lastige groepen geen specifieke acties plaatsvinden om lastiger respondenten meer deel te laten nemen aan zo'n panel (door verschillende benaderingsvormen bijvoorbeeld). Ook wordt er niet met verkorte vragenlijst gecheckt of er bij initiële wegeraars afwijkingen plaatsvinden op de doelvariabele (bijv. Voogt, 2004).

De POROSZ-onderzoekers hebben wel aandacht besteed aan mogelijke vertekeningen door het gebruik van access pools door na te gaan of de uitkomsten uit het onderzoek waarbij van een access pool gebruik is gemaakt, andere resultaten opleveren dan een face-to-facebenadering met een steekproef op basis van *random dialing* op waarbij de respondent thuis de vragen beantwoordt (Lensveldt-Mulders, 2006). Deze vergelijking is alleen gemaakt voor personen die een WAO-uitkering ontvingen.
De resultaten laten zien dat op 2 van de 8 gevoelige vragen er tussen beide steekproeven significante verschillen zijn (4 en 5 procentpunt) in de mate van melding van regelovertreding. Omdat in het ene geval de access pool hogere

prevalenties oplevert en in het andere geval de random-dialingsteekproef, is er blijkbaar geen sprake van een systematische vertekening. Dat is een argument voor het gebruik van de relatief goedkope access pools in plaats van een dure random-dialingmethode gevolgd door bezoek aan huis.
Het feit dat de uitkomsten van beide methoden niet noemenswaardig lijken te verschillen, is uiteraard een belangrijke bevinding. Het lost daarmee nog niet het probleem op dat zowel bij de face-to-face-meting als bij de access pools hetzelfde soort selectie kan optreden in de responsgroep. Het is goed voorstelbaar dat regelovertreders in dezelfde mate weigeren deel te nemen aan access pools als aan losse onderzoeken. Beide leveren dan wellicht dezelfde resultaten op maar dat is nog geen argument dat precies wordt gemeten wat wordt bedoeld.

Wat hier verder relevant is te vermelden, is dat alleen respondenten die het Nederlands goed beheersen in de steekproef zijn opgenomen. Het zal duidelijk zijn dat allochtonen daardoor zwaar ondervertegenwoordigd zijn. Het POROSZ-onderzoek en de uitkomsten daarvan hebben dus alleen betrekking op uitkeringontvangers die het Nederlands goed beheersen. Dat is in onderling overleg tussen opdrachtgever en onderzoekers overeengekomen bij de start van de metingen in 2000. Dat houdt in dat wanneer de POROSZ-resultaten worden gebruikt voor het beoordelen van de effectiviteit van beleid en de te behalen prestaties (hetgeen niet door de onderzoekers maar door SZW zelf plaatsvindt) er in feite nog een aanname geldt: namelijk dat autochtonen (Nederlandssprekenden) en allochtonen (niet-Nederlandssprekenden) een vergelijkbaar nalevingsniveau kennen (aangezien in de begroting geen verschillende prestatie-indicatoren voor naleving worden gehanteerd voor verschillende bevolkingsgroepen).[2] Dit is een aanname waar het nodige op af te dingen lijkt. Uit het evaluatieonderzoek naar het zogenaamde LAT-project waarin een drietal pilotgemeenten experimenten in het kader van hoogwaardig handhaven uitvoerden, bleek dat Nederlanders ten opzichte van niet-Nederlanders (of zij met een dubbele nationaliteit) minder vaak informatie achterhielden voor de sociale dienst, minder vaak een boete opgelegd kregen en vaker (terecht) van mening waren dat ze verplicht waren ongevraagd alle informatie op te geven aan de sociale dienst die nodig is om recht op uitkering en de hoogte daarvan te bepalen (Soethout en Van Waveren, 2003).[3] Het lijkt verstandig de gezamenlijke beslissing om alleen respondenten die het Nederlands goed beheersen op te nemen in de POROSZ-metingen, te heroverwegen. De kans groot is dat mensen die het Nederlands onvoldoende beheersen systematisch afwijken van de huidige responsgroep en daarbij vormen ze een substantiële groep van het totaal aan uitkeringontvangers.

[2] Deze aanname is eigenlijk nog breder van strekking, omdat het feitelijk gaat om taalvaardigheid. Die varieert ook binnen de groep autochtonen, bijvoorbeeld vanwege analfabetisme, maar ook vanwege gebrekkige handelingsbekwaamheid. Denk bijvoorbeeld aan dementie in geval van ouderen.

[3] Aanvullende analyses op het bestaande datamateriaal hebben deze gegevens opgeleverd.

Het opnemen van niet-Nederlandssprekenden is overigens geen eenvoudige opgave. Taalproblemen vormen een belangrijke oorzaak van het niet mee werken aan bijvoorbeeld enquêteonderzoek door het CBS. Ook moeten meer contactpogingen worden gedaan onder allochtonen om hetzelfde respons-percentage te halen en leveren verschillende dataverzamelingsmethodieken verschillen in respons op. Allochtonen beschikken veel minder vaak over een vaste telefoonlijn waardoor voor telefonische enquêtes of telefonische wervingsrondes (via random dialing) verschillen gaan ontstaan voor de kans dat allochtonen in de steekproef terechtkomen. Datzelfde geldt voor aan-sluiting op internet. Allochtone huishoudens beschikken in veel mindere mate over (breedband)internetverbindingen. Door dataverzameling alleen via het web te laten verlopen, maken allochtone groepen minder kans om te kunnen deelnemen aan het onderzoek.
Dat betekent dat het gaan opnemen van niet-Nederlandssprekenden in het POROSZ-onderzoek een ingewikkeld traject gaat worden. Er zullen, afhan-kelijk van de uitkomsten van deze studie, anderstalige vragenlijsten gemaakt moeten worden, anderstalige interviewers ingezet moeten worden, er zal misschien gebruikt gemaakt moeten worden van *multi channeling* (meer dan1 dataverzamelingstechniek). Dat betekent een grotere tijdsinvestering in het onderzoek en daarmee ook meer benodigde financiële middelen.
Desondanks is een ingewikkelder en duurder dataverzamelingsproces op zich geen steekhoudend argument om juist de niet-Nederlandssprekenden niet in een onderzoek naar regelovertreding mee te nemen (gezien de resultaten en de verwachting dat alleen al door een gebrek aan kennis regelovertredingen vaker kunnen voorkomen onder allochtonen).

Ad B1: Non-responsafwijking
Bij de non-responsafwijking speelt de vraag of groepen die niet willen mee-werken aan een onderzoek op de doelvariabele afwijken van de groep die wel wil meewerken. In dit geval is het van belang te weten of de groep uit de access pool die niet heeft gereageerd op de oproep de enquête in te vullen meer (of minder) geneigd is regels na te leven dan degenen die wel hebben gereageerd. Het onderzoek is aangekondigd bij de responsgroep als een algemeen onderzoek ‘over een heleboel zaken die voor u belangrijk zijn in verband met uw uitkering en de arbeidsmarkt’ en niet als een onderzoek naar ‘fraude met uitkeringen’. Dat maakt dat er voldoende reden is om aan te nemen dat op dat punt mensen niet bewust selectief hebben afgehaakt.
Wel blijft het in theorie mogelijk dat de vragenlijst door specifieke groepen minder is ingevuld waardoor afwijking van het nalevingspercentage kan optreden. De uitkomsten van het onderzoek laten bijvoorbeeld zien dat hoger opgeleide uitkeringsgerechtigden over het geheel meer geneigd zijn regels van de uitkering te overtreden dan lageropgeleiden. Opleidingsniveau is een achtergrondkenmerk waarop de responsgroep niet is gewogen (bij gebrek aan een goede populatieverdeling). Tegelijkertijd is bekend dat lageropgeleiden minder vaak aan vragenlijstonderzoek meewerken. Dat zou een overschatting van de niet-naleving tot gevolg kunnen hebben. Ons idee is echter dat een dergelijke selectie dan waarschijnlijk al in de fase van de steekproefsamen-

stelling is geschied en dat er van een zelfstandig effect daarna nauwelijks meer sprake zal zijn.

Ad B2: Responsafwijking
Vertekening in de respons treedt op doordat de groep die wel respondeert bedoeld of onbedoeld verkeerde informatie geeft. Eén specifiek probleem dat hierbij speelt, is dat respondenten niet naar waarheid antwoorden omdat ze zich schamen voor het antwoord, de antwoorden niet stroken met hun zelf-beeld, bang zijn dat hun antwoorden tot hun persoon herleidbaar zijn. Dat wordt wel aangeduid met gevoelige vragen (of sensitive topics).
Ook in het POROSZ gaat het om gevoelige vragen. Het eventueel toegeven van uitkeringsfraude zal voor een deel van de populatie omgeven zijn met gevoelens van angst voor eventuele ontdekking, schaamte voor het moeten bekennen van oneerlijkheid et cetera.
Het toepassen van de randomized-responseaanpak op dit onderdeel lijkt daarmee een zeer verstandige poging de te verwachten responsafwijking te lijf te gaan. Door respondenten zekerheid te bieden over de anonimiteit wordt een veilige omgeving geboden om eerlijk antwoord te geven. De dobbelsteenmethode met een forced-responsemodel dwingt respondenten bij sommige uitkomsten niet-naleving te bevestigen (los van de werkelijke situatie), bij andere niet-naleving te ontkennen (ook los van de werkelijke situatie), en bij nog andere uitkomsten de waarheid in te vullen. De onderzoekers weten afzonderlijke dobbelsteenuitkomsten niet. Omdat de kansverdeling van de dobbelsteenuitkomsten bekend zijn, kan op geaggregeerd niveau berekend worden wat de omvang van de niet-naleving is.
Is de claim dat de randomized-responseaanpak meer valide resultaten oplevert dan directe benaderingen gerechtvaardigd? Er is een overzichtsstudie uitgevoerd (door de vakgroep van de Universiteit Utrecht (UU) die ook betrokken is bij de uitvoering van het POROSZ) waarin de resultaten van 38 afzonderlijke onderzoeken worden geanalyseerd (Lensveldt-Mulders e.a., 2004).
In deze zogenaamde meta-analyse zijn twee soorten studies opgenomen. De eerste soort betreft onderzoeken (6 stuks) waarin naast uitkomsten van de randomized-responseaanpak en een of meer directe bevragingsmethoden ook werkelijke populatie-uitkomsten bekend waren (bijv. uit medische dossiers). Dat maakte het mogelijk de validiteit van een randomized-responseaanpak te vergelijken met die van directe bevragingen. Uit deze evaluatie blijkt dat de randomized-responseresultaten over het geheel genomen dichter bij bekende populatiewaarden liggen dan directe dataverzamelingstechnieken. Dit verschil blijft bestaan als er gecontroleerd wordt voor de gevoeligheid van de vraag.
In een ander deel van de onderzoeken was de randomized-responseaanpak naast een andere directe bevragingsmethodiek gebruikt zonder dat de werkelijke populatie uitkomst bekend was (32 stuks). Voor die studies zijn de schattingen van de randomized-responseaanpak en die van directe bevraging onderzocht op systematische verschillen in de rapportage van gevoelige

antwoorden. De veronderstelling daarbij is dat een hogere prevalentie duidt op een meer valide methodiek.
Uit de analyse blijkt dat de resultaten op gevoelige vragen uit randomized-responsestudies hoger liggen dan bij face-to-facebenadering en schriftelijke vragenlijsten. Verschillen met andere methodieken (waaronder telefonische enquêtes) zijn niet significant vanwege zeer beperkt aantal studies dat die beide methodieken combineert.
Verder laten resultaten zien dat het verschil tussen de randomized response en directe bevraging toeneemt met de gevoeligheid van de vraagstelling. In andere woorden: de randomized response levert meer op naarmate de vraagstelling gevoeliger wordt: er wordt dan winst in validiteit geboekt.

Een tweede belangrijke bron voor beantwoording van de vraag of de randomized response meer oplevert dan andere methodieken is het paper van Lensvelt, Hout en Van der Heijden (1994). Daarin worden vier methoden vergeleken voor het meten van regelovertreding onder WAO'ers. Er is via random dialing een groep WAO'ers benaderd met verzoek tot een face-to-face-interview. Ruim 1000 respondenten hebben daaraan meegewerkt. Bij ruim 800 daarvan is een set van zeven gevoelige vragen rondom hun WAO-uitkering gesteld met behulp van een randomized-responseaanpak. Bij de overige 200 zijn dezelfde vragen via directe bevraging aan bod gekomen. Daarnaast is via een access pool aan ruim 1100 respondenten een vragenlijst voorgelegd met ook weer dezelfde set vragen. Bijna 900 kregen de vragen via een randomized-responsedesign aangeboden. De overige 250 hebben via directe bevraging hun informatie gegeven over naleving.
Om het verschil te kunnen bepalen tussen de validiteit van de randomized response tegenover directe bevraging is het van belang de resultaten verkregen via de access pool en die via random dialing te scheiden. Immers de steekproeven voor beide aanpakken verschillen en beïnvloeden de gevonden prevalenties.
Het blijkt dat één van de vragen (heb je een gewone baan naast je uitkering?) zo weinig voorkomt dat het niet mogelijk is goed een vergelijking tussen methodieken te maken. Voor de overige zes vragen blijkt dat in een aantal gevallen de schattingen nauwelijks verschillen. Wanneer we een grens aanhouden van minimaal vier procentpunt verschil dan blijven er bij de dataverzamelingsmethode via de computer drie vragen over die een duidelijk verschil tussen de randomized response en directe bevraging laten zien en bij de face-to-face-aanpak twee vragen.[4] Bij die verschillen van minimaal vier procentpunt laat de randomized-responseaanpak hogere prevalenties van niet-naleving zien (van een verschil van 4,9 procentpunt tot een verschil van 10,4 procentpunt).

[4] We kiezen hier voor een grens van vier procent omdat SZW ooit heeft aangegeven dat de betrouwbaarheidmarges bij een nalevingspercentage maximaal plus of min vier procent zou mogen bedragen. We nemen aan dat wanneer de randomized response dus minimaal een verschil van vier procent oplevert met een directe bevraging SZW dit als een relevant gegeven beschouwt.

Ook bij de kleinere verschillen laat de randomized-responseaanpak veelal hoogste prevalenties zien. Wanneer de randomized-responseaanpak *lagere* prevalenties dan directe bevraging laten zien, zijn deze dus beperkter in omvang (minder dan 2 procentpunt verschil). We zien op het eerste gezicht ook het effect terug dat de randomized-responseaanpak meer oplevert (validiteitswinst) wanneer de vraagstellingen blijkbaar gevoeliger zijn. Over het geheel gaan relatief hoge aantallen gerapporteerd via randomized-responseaanpak (boven de 10%) gaan samen met een groter verschil met een directe bevraging.
De auteurs concluderen hetzelfde op basis van het toetsen van de verschillen (waarbij de standaardfout is meegenomen). Randomized response levert een hogere rapportage van niet-naleving op voor de meeste items. Voor de items waarbij de randomized response een lager nalevingsniveau rapporteert, zijn de verschillen niet significant (p. 18). Over het geheel genomen moet hieruit toch ook worden geconcludeerd dat de randomized response naar alle waarschijnlijkheid meer valide prevalenties oplevert dan directe bevraging.
De resultaten van de randomized-responsebevraging via de computer zijn in POROSZ-3 opnieuw geanalyseerd waarbij gebruikgemaakt is van een toegevoegd Rasch-model om een deel personen te traceren dat niet volgens de randomized-responseprocedure de vragenlijst heeft ingevuld en daarop de berekeningen van de prevalenties aan te passen (we gaan daar zodadelijk nog uitgebreid op in). De overtredingspercentages volgens de berekening met die randomized-responsemethodiek liggen systematisch hoger (2 tot 6 procentpunt) waardoor de verschillen met directe bevraging dus verder toenemen (van der Heijden, Van Gils en Laudy, 2005: p. 10).

We zijn op grond van bovenstaande van mening dat de claim gerechtvaardigd is dat over het geheel genomen een randomized-responseaanpak meer valide schatters oplevert dan een directe bevragingsmethodiek. Hoewel daarmee een randomized-responseaanpak nalevingspercentages levert die dichter in de buurt van de ware waarde liggen, is niet duidelijk hoe groot de afwijking van de vragenlijstuitkomsten van de ware uitkomsten is. Immers niet alle respondenten volgen de invulinstructies op waardoor er toch nog onderrapportage plaatsvindt, bijvoorbeeld omdat mensen niet graag gedwongen worden fraude ‘toe te geven’, voor de zekerheid toch alles ontkennen of de methodiek niet snappen.
We starten met het laatste. De randomized-responseaanpak lijkt niet voor alle respondenten even eenvoudig te begrijpen. In bijvoorbeeld een (telefonisch) onderzoek van Regioplan onder werkgevers over het in dienst hebben van illegale werknemers begreep een aanzienlijk deel van de respondenten de randomized response werkwijze niet, hoewel zij daar per brief vooraf over waren geïnformeerd. (Mosselman en Van Rij, 2005). Het door de POROSZ-onderzoekers gebruikte CARR-programma (Computer Assisted Randomized Response) is volgens eigen zeggen in een cognitieve laboratoriumstudie onderzocht en de auteurs geven aan dat er geen aanwijzingen zijn dat het programma te moeilijk te bedienen of te begrijpen is, zelfs niet voor computeranalfabeten of mensen met een laag opleidingsniveau en die soms

een min of meer geïsoleerd leven leiden (Boeije en Lensvelt-Mulders 2002, verwijzing in Lensvelt, Hout en Van der Heijden (precies jaartal onbekend) en Lensvelt-Mulders, e.a. 2003).
Hoewel de randomized-responseprocedure dan niet altijd te ingewikkeld hoeft te zijn, blijft de vraag of de randomized-responseregels netjes door iedereen worden gevolgd. Bekend is ook uit de cognitieve labaratoriumstudies dat sommige personen problemen hebben met het gedwongen toegeven van regelovertreding en regelovertreders die ook binnen de randomized-response-aanpak rapporteren alsof ze keurig alle regels naleven (Lensvelt-Mulders e.a. 2003).
Omdat uit de resultaten van POROSZ-meting 2002 duidelijk werd dat een aantal respondenten wel moest afwijken van de invulinstructies is door de onderzoekers in samenwerking met derden de randomized-responseaanpak uitgebreid met een model waarin de omvang van de groep die de randomized-responseregels niet volgt of snapt, geschat kan worden (Böckenholt en Van der Heijden, 2004). Ze maken daarbij gebruik van het feit dat er in de POROSZ-onderzoeken per uitkering een aantal vragen over min of meer hetzelfde soort overtreding voorkomt (hetgeen wordt opgenomen als een latente eigenschap via een Rasch-model). Wanneer dat wordt gecombineerd met de bekende verwachtingen over de dobbelsteenuitkomsten mogen bepaalde individuele antwoordpatronen een maximaal aantal malen voorkomen. Nu blijkt dat met name het antwoordpatroon waarbij respondenten consequent ‘nee’ rapporteren te vaak voor te komen. Op grond van de te verwachten dobbelsteenuitkomsten is bekend dat er vaker gedwongen ‘ja’ gerapporteerd zou moeten zijn. In de laatste meting wordt ook voor deze consequente nee-zeggers gecorrigeerd.
Ondanks het feit dat met deze toevoeging van een correctie voor de consequente nee-zeggers aan het onderzoek een belangrijke verbetering is aangebracht, blijft de vraag of hiermee nu de ‘ware’ uitkomst wordt gemeten. We beoordelen daarbij de aanname dat in de groep nee-zeggers niet-naleving even vaak voorkomt als in de rest van de responsgroep als meest problematische onderdeel. Het is goed voorstelbaar dat juist de overtreders geneigd zijn het zekere voor het onzekere te nemen en overal ‘nee’ op te antwoorden. De berekende cijfers presenteren daarmee dan nog steeds een onderschatting van de feitelijke niet-naleving. Tegelijkertijd geeft informatie over de omvang van die groep wel een maximum aan het percentage overtreders. Als in het extreme geval inderdaad alle ‘nee-nee-nee’-zeggers overtreders zouden zijn dan kan dat percentage bij het berekende percentage overtreding worden geteld voor een indicatie van het maximum aantal nalevers (overtreders).
De auteurs concluderen zelf met de nodige voorzichtigheid dat (p. 114, POROSZ WAO/WW/WWB, 2004) het niet waarschijnlijk is dat het bij de groep ‘nee-nee-nee’-zeggers vooral om overtreders zou gaan. Tevens geven ze aan nader onderzoek te willen verrichten via secundaire analyses naar de groep die niet volgens het randomized-responsedesign antwoordt.
Hoewel deze toevoeging door middel van het Rasch-model dus een verbetering inhoudt, is daarmee het probleem van respondenten die toch nog blijven weigeren niet-naleving toe te geven of soms wel en soms niet de

randomized-responseregels volgen, niet helemaal opgelost. Verder is het Rasch-model een belangrijke toevoeging maar levert het alleen een directe schatting op van het percentage consequente nee-zeggers bij sets van minimaal drie vragen over één bepaald nalevingsniveau. Bij twee vragen moet de onderzoeker op basis van andere informatie of veronderstellingen parameters in het model fixeren. Dat is bijvoorbeeld in POROSZ-3 gebeurd door het percentage respondenten dat niet volgens het randomized-response-design antwoordt tussen opeenvolgende metingen als constant te veronderstellen of bij een ontbrekende waarde voor een nalevingsverplichting een percentage afkomstig van een soortgelijke verplichting te gebruiken als proxy (Van der Heijden, Van Gils en Laudy, 2005: p. 116-117).

Ondanks onvolkomenheden is er voldoende reden aan te nemen dat de randomized-responseaanpak als het gaat om de validiteit superieur is aan het recht-toe-recht-aan vragen stellen aan respondenten. De gevonden percentages liggen naar alle waarschijnlijkheid dichter bij de werkelijke waarden en het draagt daarmee bij aan de validiteit van de uitkomsten.

De POROSZ-onderzoekers geven verder nog aan dat zelfs indien de huidige randomized-responseaanpak nog steeds leidt tot een onderschatting van het nalevingspercentage, de waarde van het onderzoek gelegen is in het feit dat er een trend kan worden gemeten. Dit roept vervolgens de vraag op of het POROSZ dan zou kunnen doorlopen zonder randomized-responseaanpak (de nalevingspercentages schieten dan eenmalig omhoog maar als SZW alleen nog maar in de trend is geïnteresseerd, hoeft dat geen probleem te zijn). De belangrijkste tegenwerping van de onderzoekers tegen dit argument is dat er een grote kans bestaat dat voor sommige regels dan nalevingspercentages van 100 procent gevonden gaan worden (niemand geeft meer regelovertreding toe) waardoor berekeningen over trends onmogelijk worden. Wanneer we de resultaten uit het onderzoek naar de verschillen tussen randomized-responseaanpak en directe bevraging voor de WAO bekijken dan lijkt daarvoor wel wat te zeggen. Met name bij bevraging van de access pool zonder de randomized response worden nogal wat prevalenties van lager dan één procent gevonden (3 van de 7).

Vraagstelling en definiëring naleving

De precieze definitie van de ‘mate van naleving’ (en dus ook de prestatie-indicator) ligt niet vast. De vraag is of het gaat om de groep personen die een of meer regels overtreedt ten opzichte van de totale groep die aanspraak maakt op de betreffende uitkering, of dat het gaat om de omvang van de groep overtreders ten opzichte van de subgroep die *had kunnen overtreden*. Een voorbeeld betreft de mensen in de WAO die zich nadat zij zich beter voelen, dat niet hebben gemeld bij de uitkeringsinstantie. Stel dat van de honderd WAO’ers zich er veertig beter zijn gaan voelen en dat daarvan dertig dat wel gemeld hebben. Is het niet-nalevingsniveau dan tien procent (ten opzichte van het totaal) of een kwart (ten opzichte van de personen die zich ook daadwerkelijk beter hebben gevoeld)?

De vraagstelling in het POROSZ-onderzoek is zodanig dat voor 13 van de 24 vragen bij de WAO, WW en WWB alleen de naleving ten opzichte van de totale groep uitkeringsontvangers berekend kan worden en niet voor de groepen die ook daadwerkelijk de mogelijkheid hebben gehad een regel te overtreden.
Dat heeft enerzijds met de randomized-responsemethodiek te maken. Bovengenoemde vragen betreffen eigenlijk samengestelde vragen. “Voelde u zich beter?” en “Zo ja, heeft u dat gemeld?” worden in de randomized-responseaanpak tot één vraag samengenomen: “Voelde u zich beter zonder dat u dat heeft gemeld?”. Mensen die daarop ‘nee’ antwoorden, betreffen degenen die zich niet beter hebben gevoeld èn degenen die zich wel beter hebben gevoeld en dat keurig hebben gemeld. In de randomized-response-aanpak is het niet mogelijk deze vragen los te knippen omdat de antwoordpatronen van individuele respondenten geen betekenis hebben.
Anderzijds zijn in een onderzoek waarbij gevoelige onderwerpen aan de orde komen ook dergelijke selectievragen vooraf (“Voelde u zich beter?”) al in zekere zin gevoelig en zouden dus via een randomized-responsemethodiek aangepakt moeten worden.
Dit definitievraagstuk is dus niet zozeer een oorzaak van toepassing van de randomized-responseaanpak. Wel biedt de randomized-responseaanpak minder flexibiliteit in de vraagstelling in vergelijking met een regulier zelfrapportageonderzoek. De POROSZ-onderzoekers hebben wel aangegeven mogelijkheden voor verbetering in de vragenlijst te zien door selectievragen los te knippen van vragen naar regelovertreding. Sommige selectievragen zouden aan het begin van de vragenlijst gesteld kunnen worden. Daardoor komen ze los te staan van de gevoeliger context. Pas veel later in de vragenlijst komen de vragen naar regelovertreding waarbij alleen de geselecteerde respondenten die vragen via de randomized-responseaanpak krijgen voorgelegd.

2.3.2 Betrouwbaarheid

De opdrachtgever van POROSZ heeft in de startnotitie aangegeven dat de betrouwbaarheidgrenzen van schattingen door de randomized-response-methodiek nogal ver uit elkaar liggen. Dat maakt het moeilijk om tot uitspraken over ontwikkelingen in de tijd te komen.
Het Ministerie van SZW heeft in de begroting vanaf 2005 het nalevingsniveau als prestatie-indicator opgenomen. In bijlage 2 hebben we de streefpercentages opgenomen voor de regelingen die in het kader van dit onderzoek relevant zijn. Relevant in verband met de noodzakelijke betrouwbaarheid van schatting van nalevingspercentages zijn de verschillen tussen jaren. Uit de tabel blijkt dat gemiddeld genomen er tussen jaren een stijging van het nalevingspercentage van één à twee procentpunt wordt verwacht (wordt nagestreefd).
Er wordt overigens door de POROSZ-onderzoekers op gewezen dat de aandacht voor de nalevingsindicatoren alsmede de wens om ontwikkelingen daarin goed in kaart te kunnen brengen, pas later in het traject naar voren zijn

gekomen. De oorspronkelijke afspraak was namelijk dat een nalevingspercentage van 93 procent met een betrouwbaarheidsinterval van ongeveer vier procentpunt zou moeten hebben. Daarop zijn ook de huidige responsaantallen gebaseerd.

De vraag is nu of de huidige steekproefomvang voldoende is om na te gaan of de beleidsdoelen (gemeten via prestatie-indicatoren) worden gerealiseerd. Wanneer in het onderzoek een ontwikkeling van enkele procentpunten verschil ook in de steekproef wordt waargenomen, is de vraag of dit om een 'echt' verschil gaat of dat het verschil met grote waarschijnlijkheid aan steekproef-fluctuaties toegeschreven moet worden (gaat het om een significant verschil). Voor de bruikbaarheid van het onderzoek in termen van beoordeling van de effectiviteit van het handhavingsbeleid (een van de doelstellingen van het POROSZ) betekent dat, dat bij tweejaarlijkse peilingen, gevonden verschillen van drie à vier procentpunt tussen opeenvolgende jaren niet aan random-steekproefafwijkingen toegeschreven zouden moeten kunnen worden.

Als rekenvoorbeeld nemen we hier de nalevingsmaat voor de verplichting weer werk te vinden (WW). Die ligt voor 2004 op 66 procent. Als prestatiemaat wordt een niet nader genoemde stijging voorgesteld vanaf 2005 (begroting SZW, p. 83). Laten we eens aannemen dat een stijging van anderhalf procent-punt per jaar hiervoor nagestreefd zou worden (gegeven de stijgingen bij de andere prestatie-indicatoren). Dat betekent dat tussen twee POROSZ-metingen een stijging van drie procentpunt gerealiseerd zou moeten worden. Dan moet de betrouwbaarheid van de metingen (de precisie waarmee de nalevingspercentages geschat worden) wel zodanig zijn dat een gevonden verschil van drie procentpunt tussen twee metingen ook een reëel verschil moet weerspiegelen en niet toegeschreven moet kunnen worden aan toeval (een significant verschil dus).
Op grond van de streefpercentages en de uitkomsten van vorige POROSZ-metingen kunnen we berekenen wat de steekproefomvang zou moeten zijn om een dergelijk verschil significant aan te tonen.
Wanneer we de uitkomsten van de randomized-responseberekeningen als uitgangspunt nemen (de schatting van 65,5% met betrouwbaarheidsinterval, p. 108 van het rapport over WAO, WW en WWB, met dank aan Peter van der Heijden voor zijn uitleg op dit punt) dan hebben we een responsgroep van 3500 personen nodig om een stijging van drie procentpunt significant aan te tonen.[5]
Die benodigde steekproefomvang hangt mede af het initiële nalevings-percentage. In dit rekenvoorbeeld zijn er voor een stijging van drie procentpunt van 66 naar 69 procent de grootste steekproeven nodig. Voor bijvoorbeeld het aantonen van een stijging van drie procentpunt van 81 naar 84 procent

[5] Of preciezer geformuleerd: bij een gevonden verschil van drie procentpunt de nulhypothese te verwerpen dat het nalevingsniveau op tijdstip 2 gelijk is of kleiner dan op tijdstip 1.

(nalevingspercentage opgave inkomsten uit arbeid bij de WAO) zijn steekproeven met een minimale omvang van 1400 nodig.
Wanneer we bedenken dat de responsgroep nu uit ongeveer 800 WW-respondenten bestaat dan is duidelijk dat op dit onderdeel de opzet van POROSZ-metingen niet aan de door SZW gewenste precisie kan voldoen.

Zonder de randomized-responseaanpak zouden twee responsgroepen van 1350 WW-uitkeringontvangers volstaan om een gevonden stijging van 66 naar 69 procent significant aan te tonen. Voor het aantonen van een stijging van 81 naar 84 procent kunnen we volstaan met een kleine 900. Een reguliere aanpak levert eenvoudigweg een preciezere schatting op zodat met minder onderzoekseenheden kan worden volstaan.
Het zou wel voor de hand liggen dat, omdat directe bevraging lagere overtredingspercentages laat zien, de veranderingen die per jaar worden nagestreefd door SZW dan ook in omvang kleiner worden. Bijvoorbeeld geen drie procentpunt per jaar (van 81% naar 84%) maar nog maar twee procentpunt (van 87% naar 89%). Omdat het om een kleiner verschil gaat, zijn dan per jaar ook bij directe bevraging 1400 respondenten nodig (evenveel als bij toepassing van de randomized-responseaanpak).
Hoewel de directe bevragingsmethode dus preciezere schatters oplevert, is de vraag wat deze precisie waard is als we eerder hebben geconcludeerd dat de randomized-responseaanpak meer valide is èn het de vraag is of het iets oplevert als we verschillen tussen jaren willen toetsen en we ervan uitgaan dat we bij de directe bevraging kleinere verschillen willen gaan toetsen dan bij de randomized-responseaanpak.
Om per nalevingsindicator een goed oordeel te kunnen vellen over de noodzakelijke omvang van steekproeven is hoe dan ook empirisch materiaal nodig waaruit berekend kan worden wat de verschillen in de betrouwbaarheid van de schattingen zijn tussen een randomized-responseaanpak en directe bevraging. Alleen met die gegevens kan berekend worden hoe groot de benodigde steekproeven zullen zijn voor het aantonen van veranderingen.

Verder past een opmerking over de betrouwbaarheid wanneer met selectie-vragen vooraf zou worden gewerkt. Omdat de mate van naleving dan anders wordt gedefinieerd (niet-nalevers ten opzichte van alleen hen die zouden *kunnen* overtreden) wordt ook de berekening van de betrouwbaarheids-intervallen anders. Daarin moet nu het relevante deel van responsgroep worden betrokken. Grofweg kan daarbij aangehouden worden dat wanneer de responsgroep door selectie met een factor 2 wordt verkleind het betrouwbaar-heidinterval met een factor tweemaal wortel twee wordt vergroot. Een selectie van een kwart vergroot het interval met tweemaal wortel vier, en dergelijke. Dat maakt dat met name wanneer de selecties een fors deel uitkeringsontvan-gers gaat uitsluiten er grote veranderingen zullen zijn in de betrouwbaarheids-intervallen. Dat gaat op voor zowel de randomized respons als bij een eventuele directe bevraging. Hoe de betrouwbaarheidsintervallen gaan veranderen, hangt uiteraard af van de groepen die overblijven na selectie. Gezien het feit dat het hier om een fors aantal vragen gaat waar deze selectie

in aan de orde komt, waarvan vaak ook nog onduidelijk is wat de omvang is van de selecties zal worden (hoeveel van de WAO'ers zal zich wel eens beter hebben gevoeld?), gaat het te ver hier te proberen aan te geven hoe groot de marges dan zouden kunnen worden.

2.3.3 Bereik

Het Ministerie van SZW heeft in de startnotitie een lijst met uitkeringen aangegeven waar streefniveaus voor naleving zijn of worden vastgesteld en waar vroeg of laat dus gegevens voor beschikbaar moeten zijn. Het gaat om de WWB, de WW, de WAO (vanaf januari jl. de WIA), de Anw, de AOW en de AKW.
De laatste POROSZ-meting heeft betrekking op al deze wetten. Van de WWB, de WW en de WAO/WIA ontbreekt daarin overigens nog het vaststellen van het nalevingsniveau van deelname aan een traject. Daarmee dekt het POROSZ-onderzoek momenteel dus in de laatste opzet niet alle door SZW gewenste nalevingsniveaus. Dit betekent niet dat het POROSZ ongeschikt zou zijn omdat in de toekomst wel te doen. De systematiek kan uiteraard ook uitgebreid worden naar reïntegratieverplichtingen.
Een ander onderwerp dat onder het thema 'bereik' valt, is het meten van de omvang van uitkeringsfraude in bedragen (naast de omvang als proportie van het aantal uitkeringontvangers). Hier wordt door SZW in de startnotitie niet expliciet melding van gemaakt maar het ligt voor de hand dat terugdringen van fraude ook (en misschien wel vooral) kan plaatsvinden door niet alleen te proberen het aantal overtredingen terug te dringen maar ook de bedragen die ermee zijn gemoeid. Dat maakt het dat een onderzoeksmethodiek waarin niet alleen aantallen fraudeurs maar ook het fraudebedrag meegenomen kan worden, de voorkeur verdient.
Tot nu toe zijn in de POROSZ-metingen geen fraudebedragen opgenomen. Wel is er in de laatste meting een pilot geweest waarin kwantitatieve vragen zijn opgenomen (bijv. aantal dagen zwart gewerkt). Daaruit blijkt dat het mogelijk is naar bedragen te vragen (het gaat dan om categorieën van bedragen) maar dat, net als het geval is bij de ja-nee-vragen, er respondenten zijn die de randomized-responseregels niet volgen. De auteurs hebben aangegeven dat verdere toepassing van deze uitbreiding pas mogelijk is als er een statistische correctie is ontwikkeld voor deze groep respondenten.

2.3.4 Doorlooptijd

De opdrachtgever van POROSZ heeft aangegeven dat de uitkomsten van een peiling rondom naleving over een bepaald jaar uiterlijk in februari in het volgend jaar beschikbaar zouden moeten komen. Het POROSZ-rapport over de AOW, AKW en Anw over 2004 is gepubliceerd in november 2005. De elektronische vragenlijsten (waarin respondenten wordt gevraagd een jaar terug te kijken) zijn in de maanden september en oktober 2004 ingevuld (voor het WAO/WW en WWB-deel van POROSZ-2004 zijn dergelijke gegevens niet uit de rapportage te halen). Dat betekent dat hier tussen peildatum en

rapportage een jaar is verstreken. Inmiddels is een aantal elementen binnen POROSZ zo gestandaardiseerd (bijvoorbeeld het BH-model) dat de onderzoekers in staat zijn de specifieke nalevingspercentages eerder op te leveren (voor eind februari) en dat de verdere analyses en uitsplitsingen later volgen.

2.3.5 Kosten

De kosten die gemoeid zijn met randomized-responseonderzoek zijn hoger dan wanneer er sprake is van rechtstreekse bevraging. Dit wordt met name veroorzaakt door het grotere aantal respondenten dat benaderd moet worden om tot acceptabele betrouwbaarheidsmarges te komen.

2.3.6 Draagvlak in veld

Draagvlak voor een zelfrapportageonderzoek zoals het POROSZ lijkt bij de meeste uitkeringsinstanties aanwezig te zijn. Bij de SVB wordt echter weinig meerwaarde van POROSZ verwacht omdat de grootste risico's die aan de volksverzekeringen verbonden zijn, goed te beheersen zijn met behulp van administratieve controles en huisbezoeken.

2.3.7 Privacybescherming

Bij het zelfrapportageonderzoek spelen momenteel in de POROSZ-opzet geen problemen met betrekking tot privacy van respondenten. Dat wordt anders wanneer van uitkeringsadministraties gebruikgemaakt gaat worden.
De ervaring van Regioplan leert dat door goed overleg en uitgebreide waarborgen het vrijwel altijd mogelijk is beoogde respondenten voor onderzoek te werven.

2.4 Samenvattend overzicht

We geven in deze paragraaf een samenvattend overzicht van de bevindingen met betrekking tot de criteria.

Validiteit
De steekproef van het POROSZ-onderzoek in de opzet van 2004 kent twee belangrijke beperkingen.
Door het gebruik van de access pools is goed voorstelbaar dat door het vrijwillige karakter van het onderzoek relevante groepen regelovertreders niet in beeld komen. Door de onduidelijke steekproefsamenstelling is nauwkeurige responsanalyse niet mogelijk en kan ook niet goed aangegeven worden of en in welke mate de responsgroep een goede afspiegeling vormt van de populatie. Een steekproef via de Belastingdienst, het UWV en de SVB verdient daarom de aanbeveling.

Door het niet opnemen van uitkeringsontvangers die geen Nederlands spreken wordt momenteel de mate van regelovertreding naar alle waarschijnlijkheid onderschat. Ondanks problemen die het oplevert deze groepen in de steekproef op te nemen, is het uit het oogpunt van de validiteit van het onderzoek daar in de toekomst aandacht aan te besteden.

Het toepassen van de randomized-responseaanpak levert (ondanks alle bezwaren die er tegen in te brengen zijn) over het algemeen hogere prevalenties voor regelsovertreding. Dat interpreteren we als meer valide schattingen van de ware uitkomsten. Vermoedelijk is er nog steeds sprake van een zekere mate van onderschatting ondanks toevoeging van het Rasch-model maar die wordt begrensd door het percentage respondenten dat niet volgens de randomized-responsemethodiek antwoordt. Verder is de validiteitswinst groter naarmate de gevoeligheid van de vraagstelling toeneemt.
Wanneer we aannemen dat de gevoeligheid van vragen naar regelovertreding niet sterk zullen wijzigen in de loop van de tijd valt te overwegen een randomized-responseaanpak met een directe bevraging te combineren. Bijvoorbeeld tweejaarlijks een directe bevraging die eens in de drie metingen wordt uitgebreid met een randomized-responseaanpak om de verschillen te blijven controleren.

De mate waarop niet-naleving wordt gedefinieerd zou bepalend moeten zijn voor de wijze waarop meting moet plaatsvinden. Momenteel is de wijze van vraagstelling bepalend voor de definitie. In een groot aantal vragen is een voorwaardelijke conditie opgenomen en wordt gevraagd naar eventueel overtreden. Respondenten die niet aan de conditie voldoen, worden eveneens tot de nalevers gerekend. Dat betekent dat het nalevingspercentage ook wordt beïnvloed door de uitkeringsontvangers die niet de mogelijkheid hebben tot niet-naleving. Een duidelijke beslissing van de opdrachtgever op dit onderdeel is gewenst om het onderzoek meer richting te kunnen geven. De randomized-responseaanpak heeft ten opzichte van een directe bevraging als nadeel dat het ingewikkelder is om de voorwaarden helemaal los te maken van de vraag naar overtreding (en dus relevante selecties te maken).

Betrouwbaarheid
De huidige aantallen respondenten zijn onvoldoende om de stijgingen in nalevingsniveaus die SZW in de begroting vastlegt significant te kunnen aantonen. Het is afhankelijk van de gevonden nalevingsniveau en spreidingen hoeveel extra respondenten nodig zijn. In een randomized-response-methodiek zijn er afhankelijk van het nalevingsniveau per jaar ongeveer tussen de 600 en 2700 extra respondenten nodig om een stijging van drie procentpunt te kunnen aantonen. Bij directe bevraging ligt dat tussen de 100 en 550.
Wanneer we ook rekening houden met lagere prevalenties bij directe bevraging, waardoor het voor de hand ligt kleinere veranderingen tussen jaren na te streven, kàn het zijn dat bij directe bevraging even grote steekproeven

nodig zijn als bij een randomized-responseaanpak (mogelijk zelfs nog groter). Alleen empirische gegevens kunnen hierover uitsluitsel geven.
Wanneer de definitie van niet-naleving op relevante selecties van uitkering-ontvangers wordt gebaseerd en de huidige betrouwbaarheidseisen blijven gelden dan is het aantal benodigde respondenten nog groter (afhankelijk van de omvang van de selecties).

Bereik
Het is mogelijk om alle nalevingniveaus via zelfrapportage in kaart te brengen. Te overwegen valt om een deel van de nalevingniveaus niet via de randomized response in kaart te brengen omdat de omvang van de niet-naleving dermate beperkt is dat validiteitswinst van de randomized-response-aanpak minimaal is.

Doorlooptijd
De doorlooptijd van het POROSZ in de huidige vorm biedt de mogelijkheid om op het door SZW gewenste tijdstip (februari t+1) resultaten over de mate van niet-naleving in het meetjaar (t) gereed te hebben. Voor alle duidelijkheid melden we dat een andere steekproef en of andere dataverzamelingsmethode de looptijd van het onderzoek (fors) kunnen verlengen.

Kosten
De kosten voor randomized-responseonderzoek zijn hoger dan de kosten die gemoeid zijn met rechtstreekse bevraging van cliënten.

Draagvlak in veld
Het draagvlak in het veld voor een zelfrapportageonderzoek is grotendeels aanwezig, met uitzondering van de SVB.

Privacy
Bij het zelfrapportageonderzoek spelen momenteel in de POROSZ-opzet geen problemen met betrekking tot privacy van respondenten. Dat wordt anders wanneer van uitkeringsadministraties gebruikgemaakt gaat worden.
De ervaring van Regioplan leert dat door goed overleg en uitgebreide waarborgen het vrijwel altijd mogelijk is beoogde respondenten voor onderzoek te werven.

3 ASELECTE CONTROLES EN DCE

3.1 Inleiding

Door middel van aselecte controles kan de mate van regelovertreding in kaart gebracht worden. Een willekeurige steekproef van de doelgroep van potentiële regelovertreders wordt onderworpen aan een controle. Deze controle kan redelijk 'licht' zijn, bijvoorbeeld door alleen administraties en dossiers te onderzoeken. Controles kunnen echter ook diepgaander worden uitgevoerd, door bijvoorbeeld de potentiële regelovertreder op te roepen voor een gesprek of door 'waarnemingen ter plaatse' uit te voeren.
We beschrijven de verschillende soorten controles en geven vervolgens in paragraaf 3.3 enkele praktijkvoorbeelden van het gebruik van (aselecte) controles. Aan het gebruik van aselecte controles voor het vaststellen van de mate van regelovertreding zijn een aantal voor- en nadelen verbonden. We zetten ze in paragraaf 3.4 op een rij aan de hand van de criteria validiteit, betrouwbaarheid, bereik, doorlooptijd, kosten, draagvlak en privacybescherming.
Aanvullend geven we kort de werkwijze van Detection Controlled Estimation (DCE) weer en bespreken we de mogelijkheid om dit bij onderzoek naar regelovertreding in de sociale zekerheid als methode in te zetten. We besluiten het hoofdstuk met een korte samenvatting.

De uitkomsten in dit hoofdstuk zijn gebaseerd op literatuuronderzoek, onderzoeksresultaten van uitkeringsinstanties en interviews met deskundigen.

3.2 Controles

Doel van controles
Uitkeringsinstanties kunnen op verschillende manieren en met verschillende doelen controles uitvoeren. Het doel van een controle kan zijn om vast te stellen of een uitkering *rechtmatig* wordt verstrekt. Er wordt dan gekeken of de hoogte van de uitkering overeenkomt met het recht van de cliënt. Om dat recht vast te stellen, kan bijvoorbeeld worden gekeken naar het inkomen, het eventuele vermogen en de woon- en huishoudsituatie van de cliënt.[1]
Een controle kan ook tot doel hebben om te zien of een uitkering *doelmatig* wordt verstrekt. Dan wordt gekeken of een cliënt voldoet aan zijn eventuele verplichting om zelfstandig aan het werk te gaan, mee te werken aan trajecten naar werk of activering of mee te werken aan herstel.

[1] De relevante onderwerpen om te onderzoeken bij het vaststellen van het recht op uitkering, verschilt per type uitkering.

Aard van controles
Een groot deel van de controles die door uitkeringsinstanties worden uitgevoerd, vallen onder de noemer *'reguliere controles'*. Dit zijn controles die standaard worden uitgevoerd voor alle cliënten, bijvoorbeeld bij een aanvraag van een WWB-uitkering of bij een periodieke controle van het (gehele) cliëntenbestand. Reguliere controles worden doorgaans uitgevoerd door 'gewone' klantmanagers/medewerkers van de uitkeringsinstantie.
Wanneer er naar aanleiding van en vermoeden van fraude, aan de hand van een risicoprofiel of bijvoorbeeld in het kader van een specifiek project controles worden uitgevoerd bij een selectie van cliënten, is sprake van *bijzonder onderzoek*. Dit wordt doorgaans uitgevoerd door speciaal daarvoor opgeleide handhavingsmedewerkers.
Wanneer tijdens een regulier of bijzonder onderzoek blijkt dat de omvang van de vermoede fraude zo hoog is dat het de grens van € 6.000,- overschrijdt en daarmee aangemeld dient te worden bij het Openbaar Ministerie, is sprake van een *opsporingsonderzoek*, uitgevoerd door ambtenaren met opsporings-bevoegdheid.

Controlemethodieken
Controles kunnen op verschillende manieren uitgevoerd worden. De voor de cliënt minst ingrijpende variant bestaat uit *administratief* onderzoek waarbij gegevens uit dossiers of registraties worden vergeleken met andere administratieve bronnen (registraties, bewijsstukken et cetera).
Vervolgens kan (vaak volgend op administratief onderzoek) de cliënt *persoonlijk* bevraagd worden over zijn situatie. Dit kan bijvoorbeeld plaatsvinden in de spreekkamer bij de uitkeringsinstantie.
Een intensievere methode voor controle is het *huisbezoek* waarbij medewerkers van de uitkeringsinstantie verwacht of onverwacht bij de cliënt thuis komen. Bij huisbezoeken wordt in de eerste plaats gekeken naar de woonsituatie en huishoudsamenstelling van de cliënt. Eventuele zaken die kunnen duiden op het niet-naleven van andere verplichtingen, kunnen ook in een huisbezoek aan het licht komen.
Wanneer er een concreet vermoeden van fraude bestaat, kunnen *waarnemingen ter plaatse* uitgevoerd worden. Dit betekent dat de gangen van de cliënt nagegaan worden, door bijvoorbeeld te 'posten' bij het huis van de cliënt, of hem te volgen wanneer hij naar een (vermoed) zwartwerkadres gaat.

Selecte of aselecte controles
De hiervoor beschreven controlemethodieken kunnen toegepast worden op een selecte groep cliënten, waarbij de selectie bijvoorbeeld gemaakt wordt aan de hand van risicoprofielen of signalen van onrechtmatig gedrag (bijvoorbeeld opgemerkt door een fraudealerte medewerker of na een tip).
De controles kunnen ook uitgevoerd worden bij een willekeurige groep cliënten (aselect).
Wanneer de groep cliënten die aselect gecontroleerd wordt een representatieve afspiegeling vormt van de totale cliëntenpopulatie, betekent dit dat de uitkomsten van de controles gelden voor de totale populatie. Bij het

risicogestuurde controles, zijn de uitkomsten niet representatief voor de totale populatie. Op basis van de kenmerken van de onderzochte cliënten en de kenmerken van de populatie dient eerst herweging van de controle-uitkomsten plaats te vinden voordat uitspraken voor de gehele populatie kunnen worden gedaan.

Controlekans en detectiekans
Wanneer de mate van regelovertreding wordt bepaald aan de hand van controles, spelen de begrippen *controlekans* en *detectiekans* een rol. Niet iedere cliënt wordt immers (intensief) gecontroleerd en niet alle regelovertreding wordt bij controles opgemerkt.
De controlekans is de kans dat een cliënt gecontroleerd wordt, en heeft te maken met het aantal potentiële regelovertreders, de intensiteit waarmee zij gecontroleerd worden en de manier waarop gecontroleerd wordt (select of aselect). Als er nauwelijks controles plaatsvinden, is de controlekans laag. Wanneer alle potentiële regelovertreders gecontroleerd worden, is de controlekans honderd procent. Bij aselecte controles is de controlekans voor alle cliënten gelijk, bij risicogestuurde controles is de controlekans voor cliënten met een hoog risico hoger dan voor cliënten bij wie een laag risico op regelovertreding vermoed wordt.
De detectiekans is de kans dat een overtreding ook daadwerkelijk opgemerkt wordt bij een controle. Dit verschilt per verplichting, aangezien de ene vorm van regelovertreding veel eenvoudiger vast te stellen is dan de andere. De detectiekans kan echter ook samenhangen met kenmerken van de controles (bij intensievere controles is de detectiekans groter) of met de bekwaamheid van de controleurs.

Controle als momentopname
Een controle is een *momentopname*: een eventuele regelovertreding kan op het moment van controle gedetecteerd worden. Dit verschilt van de mogelijkheid om via persoonsonderzoek regelovertreding aan het licht te brengen. Dan kan bijvoorbeeld gevraagd worden of de cliënt in de afgelopen twaalf maanden zich niet aan de regels gehouden heeft. De regelovertreding heeft dan betrekking op een *periode* en kan daarmee hoger uitkomen dan de gedetecteerde overtredingen op een bepaald moment.

3.3 Ervaringen in de praktijk

In het onderzoek hebben wij bij verschillende uitkeringsinstanties (gemeenten, het UWV en de SVB) ervaringen opgetekend met het doen van aselecte controles.[2] Hoewel uitkeringsinstanties in de regel risicogestuurd te werk gaan om zo hun handhavingscapaciteit zo efficiënt mogelijk in te zetten, worden aselecte controles bewust gebruikt om te testen in hoeverre de risicogestuurde

[2] Niet-openbare resultaten van de controleonderzoeken kunnen wij niet in deze rapportage opnemen.

werkwijze voldoet. Ook worden aselecte controles gebruikt om risicomodellen in te richten en om de mate van naleving vast te stellen.

Controles om de mate van naleving vast te stellen
Om na te gaan in hoeverre de huishoudsamenstelling en de woonsituatie van cliënten overeenkomen met de opgegeven situatie, voert de SVB jaarlijks circa 1000 aselecte huisbezoeken uit. In theorie wil de uitkeringsinstantie langs deze weg tot risicoprofielen komen, maar de aantallen geconstateerde onrechtmatigheden zijn dermate klein dat dit niet mogelijk is.
De SVB ziet deze werkwijze als een goede manier om te peilen in hoeverre cliënten juiste informatie verstrekken over hun leefsituatie.

Controles om risicoprofielen te ontwikkelen
Door eerst aselecte controles uit te voeren, hieruit grove risicoprofielen af te leiden en deze stapsgewijs verder te testen in aselecte en selecte controles heeft gemeente X risicoprofielen opgesteld. Aan de hand van deze risicoprofielen is vervolgens het halve cliëntenbestand van de gemeente gecontroleerd (17.000 controles in totaal).
De controles bestonden uit bestandskoppelingen van de eigen uitkeringsadministratie aan eigen registraties over fraude en re-integratie, het Inlichtingenbureau, de Gemeentelijke Basis Administratie, de Rijksdienst voor het Wegverkeer (RDW), de Kamer van Koophandel, het Kadaster en een registratie waarin buurtkenmerken zijn opgenomen. Vervolgens zijn de dossiers onderzocht en zijn huisbezoeken uitgevoerd.
De controles hebben ertoe geleid dat in achttien procent van de onderzochte gevallen de uitkering is beëindigd.[3] In de meeste gevallen waarin onrechtmatigheden zijn geconstateerd, ging het om adresfraude. De uitvoerders geven aan dat deze fraudevorm het onderwerp van onderzoek was, maar dat andere fraudevormen zijn betrokken in de controles wanneer daar aanleiding toe was (bijvoorbeeld na een signaal).
Het project waarin de controles zijn uitgevoerd, was arbeidsintensief. De uitvoering was in handen van de klantmanagers en 75 nieuw geworven handhavingsspecialisten. De huisbezoeken werden uitgevoerd in teams van twee: één klantmanager en één handhavingsspecialist.
De uitkomsten van dit grootschalige project worden gebruikt om in de toekomst alle controles aan de hand van risicoprofielen uit te voeren. Om de profielen te kunnen toetsen, zullen periodiek ook aselecte controles worden uitgevoerd.

Controles om risicoprofielen te valideren
Het UWV maakt bij haar controles gebruik van risicoprofielen en heeft een deel van de controles niet aan de hand van deze profielen, maar aselect uitgevoerd. Het betrof controles op diverse verplichtingen, waarbij gebruik werd gemaakt van een algemeen risicoprofiel (dus niet toegespitst op

[3] In een deel van deze gevallen is de uitkering na stopzetting weer (gewijzigd) voortgezet.

specifieke verplichtingen). De controles bestonden uit administratieve controles, gevolgd door een huisbezoek.
Uit de aselecte controles bleek dat een aantal kenmerken op basis waarvan het risicoprofiel was samengesteld, inderdaad van invloed was op het al dan niet naleven. Het UWV voert jaarlijks deze 'check' op het bestaande risicoprofiel uit en verkent met behulp van de aselecte controles de mogelijkheden om nieuwe kenmerken aan de profielen toe te voegen.

Controles om het handhavingssysteem te valideren
Gemeente Y heeft in de afgelopen jaren haar handhavingssysteem gewijzigd en signaal- en risicogestuurd ingericht. Het gevolg is dat niet alle cliënten met enige regelmaat gecontroleerd worden. Om te beoordelen of de nieuwe handhavingssystematiek voldoet, is een aselecte steekproef van cliënten onderzocht op identiteit, woon-/leefsituatie, inkomen, vermogen en arbeidsverplichtingen. Eerst is een administratief onderzoek uitgevoerd. Daarbij is gekeken naar de volgende zaken:
- Zijn in het dossier de voor de bepaling van de rechtmatigheid van de uitkering de benodigde gegevens aanwezig en juist geïnterpreteerd?
- Is de in het dossier aanwezige informatie juist in het geautomatiseerde uitkeringssysteem ingebracht?
- Validering van de aanwezige gegevens met de actuele gegevens uit GBA, SUWI-net, IB en RDW.

In een deel van de gevallen is na het administratief onderzoek de uitkering aangepast of stopgezet. In de meeste gevallen volgde een vervolgonderzoek bestaande uit schriftelijke uitvraag van gegevens bij de klant, een persoonlijk gesprek en (aangekondigd) huisbezoek bij de klant.

Het onderzoek had tot doel om na te gaan of de rechtmatigheid van het handhavingsmodel voldoende geborgd is, en om te komen tot voorstellen voor verbeteringen in de uitvoering en in de bevordering van de nalevingsbereidheid van cliënten.

3.4 Bruikbaarheid aselecte controles

Aan de hand van een aantal criteria beschrijven wij de bruikbaarheid van aselecte controles bij het vaststellen van nalevingsniveaus in de sociale zekerheid.

Validiteit uitkomsten
In een vergelijkend onderzoek naar schattingsmethoden om de omvang van fraude te bepalen, constateren Lensvelt-Mulders e.a.[4] dat de uitkomsten van aselecte controles in de regel als 'hard' beschouwd worden. Een controle wijst

[4] Lensvelt-Mulders, G.J.L.M., G.H.C. van Gils en P.G.M. van der Heijden*, Een overzicht van schattingsmethoden voor de omvang van fraude*. Utrecht, december 2000.

uit of er wel of geen sprake is van een overtreding. Bovendien is er geen sprake van non-respons of selectiviteit van de respons, dus de controles kunnen uitgevoerd worden bij alle relevante cliëntgroepen.
Daarbij plaatsen zij echter de belangrijke kanttekening dat de validiteit van de waarnemingen van individuele controleurs niet zonder problemen is. De controleurs geven een subjectief oordeel over het al dan niet aantreffen van regelovertreding. Met name wanneer het overtredingen betreft die niet eenvoudig objectief zijn vast te stellen, vormt dit een risico voor de validiteit van de uitkomsten. De controles worden uitgevoerd door een aanzienlijk aantal controleurs en de beoordelingscriteria zullen zich niet altijd volledig laten objectiveren en expliciteren.
Daar komt bij dat de controleurs tijdens de controle de afweging kunnen maken hoe intensief zij controleren totdat regelovertreding wordt vastgesteld of uitgesloten. Dit leidt tot onbedoelde selectiviteit in de aselecte steekproef.

Al met al vormt de validiteit van de waarnemingen bij aselecte controles een risico. Hierbij past de kanttekening dat dit risico beter te beheersen is, naarmate de controles strakker te protocolleren zijn. De mogelijkheden hiertoe hangen deels samen met het te onderzoeken fenomeen (bijvoorbeeld de duidelijkheid van het begrip, de bewijsbaarheid van overtreding). Deels zijn zij afhankelijk van de institutionele omgevingen: binnen een gecentraliseerde organisatie als het UWV of de SVB is protocolleren en bevorderen van een gelijke uitvoering gemakkelijker dan in de gedecentraliseerde context waarbinnen de Bijstandswet wordt uitgevoerd.

Betrouwbaarheid uitkomsten
Wanneer de steekproef voldoende groot is, en het mogelijk is om objectief te beoordelen of wel of niet aan een verplichting voldaan is, zijn aselecte controles bij uitstek geschikt om tot betrouwbare uitspraken te komen. De betrouwbaarheid van de uitkomsten kan vergroot worden door de steekproef te vergroten.
In vergelijking met andere schattingsmethoden onderscheiden aselecte controles zich in positieve zin als het gaat om betrouwbaarheid.

Bereik
De praktijkervaringen van de uitkeringsinstanties leren dat met name overtredingen die redelijk gemakkelijk objectief vast te stellen zijn (adresfraude) goed door middel van controles achterhaald kunnen worden. Bij minder gemakkelijk objectiveerbare regelovertredingen (onvoldoende meewerken aan re-integratie, zwartwerk), lijken controles een minder bruikbaar instrument.

Doorlooptijd
Uitvoerders geven aan dat met het goed uitvoeren van controles een niet te lange doorlooptijd gemoeid mag zijn. Dit om te voorkomen dat de controles bekendheid krijgen onder de cliënten en zo de uitkomsten beïnvloeden zodat geen zuiver beeld verkregen wordt van de mate van naleving. Dit betekent dat

de gehele te controleren groep binnen een beperkte periode onderzocht moet worden, iets wat veel menskracht vergt en dus hoge kosten met zich meebrengt en het risico van verschillen in werkwijze tussen controleurs vergroot.
Een voordeel is wel dat het, met de nodige inspanningen, mogelijk is om de benodigde informatie over verslagjaar t beschikbaar te hebben in februari van het jaar t + 1.

Kosten
Wanneer aselecte controles verder gaan dan een administratieve controle, is het een kostbare methodiek. Er zijn aanzienlijke aantallen controles nodig, die bovendien arbeidsintensief zijn. Bij bepaalde verplichtingen vergt het bijvoorbeeld intensief onderzoek om vast te kunnen stellen of er sprake is van een overtreding.
Ook zijn er 'sociale kosten'[5] voor de doelgroep van potentiële regelovertreders. Wanneer de controles intensief zijn en gesprekken, waarnemingen ter plaatse of zelfs 'hinderlijk volgen' omvatten, kan dit door de doelgroep als zeer belastend worden ervaren.

Draagvlak in het veld
De combinatie van de uitvoerbaarheid van de aselecte controles en het rendement dat het oplevert voor de organisatie die de controle uitvoert, maakt dat het draagvlak voor aselecte controles in het veld niet voor iedere verplichting even groot is. Door middel van een kosten-batenanalyse kan worden bepaald dat aselecte controles niet zinvol zijn. Dit kan bijvoorbeeld omdat een overtreding nauwelijks is vast te stellen (zwartwerk) of omdat het financiële risico dat de uitvoeringsorganisatie loopt maar van korte duur is (inkomsten van partner bij AOW, inkomsten van inwonende kinderen ouder dan 16 jaar bij AKW).
Daar staat tegenover dat bij andere vormen van regelovertreding controles als effectief beschouwd worden (adresfraude, vaststellen gezamenlijke huishouding). Hierbij geldt overigens dat de uitvoeringsorganisaties soms de neiging hebben om regelovertreding zo gericht mogelijk op te sporen, en niet in eerste instantie geïnteresseerd zijn in de mate van naleving door de totale populatie. Dit zorgt ervoor dat uitvoeringsorganisaties zo veel mogelijk op zoek zijn naar valide risicoprofielen die het mogelijk maken om daar te controleren waar de hoogste pakkans bestaat. Bij aselecte controles is de pakkans kleiner, maar ontstaat direct inzicht in de mate van naleving door de totale populatie. Afgezien van het argument van de efficiënte inzet van mensen en middelen, worden aselecte controles door uitvoeringsorganisaties wel gezien als een goede methodiek om nalevingsniveaus in kaart te brengen, maar ook om risicoprofielen te valideren en handhavingssystemen te 'borgen'. In paragraaf 3.4 beschrijven we enkele praktijkvoorbeelden.

[5] Lensvelt-Mulders, G.J.L.M., G.H.C. van Gils en P.G.M. van der Heijden, *Een overzicht van schattingsmethoden voor de omvang van fraude*. Utrecht, december 2000.

Privacybescherming
Bij aselecte controles moet rekening gehouden worden met de Wet bescherming persoonsgegevens. Deze stelt dat sprake moet zijn van proportionaliteit. Dit betekent dat niet zomaar een intensieve controle op een groep cliënten plaats mag vinden terwijl er geen sprake is van een daadwerkelijke verdenking. Ervaringen in de praktijk leren echter dat wanneer een uitkeringsinstantie aan kan tonen dat bij het niet uitvoeren van de controle (omvangrijke) financiële of maatschappelijke schade ontstaat, er doorgaans toch mogelijkheden zijn om de controles uit te voeren, bijvoorbeeld binnen de kaders van een (tijdgebonden) project.

3.5 Detection Controlled Estimation

Methode
Detection Controlled Estimation (DCE) is een statistische techniek die informatie afkomstig uit controles kan verbeteren. De (waarschijnlijk terechte) veronderstelling is dat controles niet altijd overtredingen aan het licht brengen terwijl die er wél zijn (detectieprobleem). De essentie van DCE is dat controle-uitkomsten worden gerelateerd aan kenmerken van de gecontroleerde gevallen of personen enerzijds en aan kenmerken van controleurs en controleomstandigheden anderzijds. Er worden op die manier per controle uitkomst twee vergelijkingen opgesteld. In de eerste vergelijking (de overtredingsvergelijking) worden de overtredingen geschat op basis van kenmerken van de gecontroleerde. In een tweede vergelijking (de detectievergelijking) wordt de detectie geschat op basis van kenmerken van controleur en controleomstandigheden. De parameterwaarden (effecten) voor de kenmerken in de overtredingsvergelijking worden op deze wijze beter geschat dan wanneer alleen een overtredingsvergelijking wordt gebruikt. Dat laat Feinstein bijvoorbeeld in zijn studie zien waarmee hij de techniek voor DCE introduceerde (Feinstein, 1990) maar ook Jablonowksi die het heeft toegepast op meldingen van veiligheidsincidenten bij gasboringen (Jablonowski, 2002). In beide gevallen worden de effecten van de voorspellende waarde van kenmerken van de ‘gecontroleerde’ op overtredingen veranderd door de introductie van de detectievergelijking. Door het gebruik maken van de variatie in controleurs/controleomstandigheden worden er betere effecten van de kenmerken van gecontroleerden geschat en komt informatie beschikbaar over controleurs.
Er kunnen ook betere schatters worden gemaakt van het vóórkomen van niet-naleving en er komt informatie beschikbaar over welk soort controleurs en welke omstandigheden de detectiekansen vergroten (hetgeen als managementinformatie voor inspectiedoeleinden erg relevant kan zijn).

Mogelijke werkwijze schatten niet-naleving in de sociale zekerheid
In het kader van deze validatiestudie is uiteraard vooral de schatting van de niet-gedetecteerde niet-naleving relevant. Als het totale model is geschat (de

overtreding en detectievergelijking) kan afgeleid worden wat de omvang van de niet-naleving is die niet door detectie aan het licht kon worden gebracht. Wat daarvoor dus in elk geval nodig is, is een model waarmee de kans op niet naleving aan de hand van kenmerken van in dit geval een uitkeringsontvanger geschat kan worden (met behulp van de overtredingsvergelijking). Dat kunnen achtergrondkenmerken zijn (leeftijd, uitkeringsduur, geslacht, woonplaats, et cetera) maar ook zaken als motivatie, werkervaring, ziekteverloop, et cetera. Omdat er bij controles in de regel geen vragenlijsten worden afgenomen, zal een en ander toch beperkt worden tot formele achtergrondkenmerken. Die zullen ongetwijfeld eenvoudig beschikbaar kunnen komen. De vraag is evenwel welke voorspellende waarde die hebben.

Daarnaast zijn kenmerken van de controle nodig (controleur, bevoegdheid, tijdstip controle, intensiteit van controle) voor het opstellen van de bijbehorende detectievergelijking. Ook deze kenmerken zijn waarschijnlijk beschikbaar of indien noodzakelijk eenvoudig bij te houden.
Vervolgens worden de vergelijkingen geschat via een statistische procedure waarmee de effecten van controlekenmerken op de kans op detectie en de effecten van de kenmerken van uitkeringontvangers op de kans op niet-naleving worden verkregen.

Voor het goed kunnen schatten van de kans op niet-gedetecteerde niet-naleving moet het totale model (beide vergelijkingen) ‘identificeerbaar’ zijn. Dat betekent eenvoudig gezegd dat er voldoende variatie in soort uitkeringsontvangers per soort controles moet zijn en voldoende variatie in soort controleur per soort uitkeringsontvanger. Dat is nodig om te kunnen bepalen of bijvoorbeeld een laag *geobserveerd* niveau van niet-naleving bij een combinatie van soort controleur en soort uitkeringontvanger vooral wordt veroorzaakt door een lage detectiekans of door een daadwerkelijk laag niveau van niet-naleving.
Stel dat de geobserveerde niet-naleving bij oudere mannen in de G4 erg hoog ligt (groep A). Leeftijd, geslacht en stad/platteland zijn kenmerken die we willen meenemen in de overtredingsvergelijking. Wanneer de controles in de G4 zich kenmerken door een intensieve controle (administratieve controle vooraf plus huisbezoek op meerdere momenten) dan is nodig dat we ook groepen in de totale controlegroep hebben (B en C) ter vergelijking. Het is namelijk nodig dat een vergelijkbare groep B (oudere mannen in de G4) aan een ander soort controles wordt onderworpen (alleen administratieve controle zonder huisbezoek). Ook is het nodig dat groep C, die juist op de achtergrondkenmerken verschilt (oude mannen op het platteland), dezelfde controle krijgt.

Op het moment dat soorten controles en soorten uitkeringsontvangers met elkaar samengaan, wordt zogenaamde ‘identificatie’ van het model een probleem. Al dat het geval is kan de schattingsprocedure niet goed onderscheid maken tussen detectie en overtredingseffecten en kan het model niet geschat worden. Feinstein wijst er dan ook op dat de schatting van de

niet-geobserveerde niet-naleving met de nodige voorzichtigheid geïnterpreteerd moet worden (Feinstein, 1990: p. 236). Verder geldt dat het om aselecte controles moet gaan.

Bruikbaarheid
DCE is vooral een verbetering van de informatie die uit controles wordt verkregen en niet zozeer een zelfstandige techniek. Omdat er nog geen pogingen zijn gedaan de techniek in Nederland op overtredingen in de sociale zekerheid toe te passen is door ons niet aan te geven hoe bruikbaar een en ander zou kunnen zijn.

Vragen zijn bijvoorbeeld in hoeverre uitkomsten van aselecte controles beschikbaar zijn, welke variabelen in de overtredingsvergelijkingen en detectievergelijkingen van belang zullen blijken, of de schatting van niet-geobserveerde niet-naleving (erg) afhankelijk is van de effectschattingen voor kenmerken van uitkeringsontvangers en controles, of er identificatieproblemen zullen gaan ontstaan.
Er kan in elk geval niet gebruik gemaakt worden van standaard statistische programmatuur (SYSTAT, SPSS, STATA). Dit betekent dat er voor berekeningen contact gezocht moet worden met instituten die inmiddels programmatuur hebben ontwikkeld of dat er zelf geprogrammeerd moet worden.

Dat betekent dat de meerwaarde van DCE pas kan bepaald kan worden na een controleronde waaruit gegevens beschikbaar komen over niet-naleving met betrekking tot één of meerdere verplichtingen. Alleen met echte gegevens uit een controleronde kan worden bekeken wat de relevante variabelen in de overtreding- en detectievergelijkingen zullen blijken, of dat een identificeerbaar model oplevert, wat de betrouwbaarheidmarges rondom parameterschatting voor niet-naleving zullen zijn, et cetera.

3.6 Samenvattend overzicht

In deze paragraaf geven we een samenvattend overzicht van de verschillende criteria waarmee we de bruikbaarheid van aselecte controles beoordelen en de consequenties die dit heeft per verplichting. Ook geven we kort weer in hoeverre en onder welke voorwaarden DCE van nut kan zijn.

3.6.1 Criteria

Validiteit
Bij controles kunnen regelovertredingen over het hoofd worden gezien. De mate waarin dit gebeurt, kan afhankelijk zijn van het soort overtreding, van de beoordeling van de controleur en van de intensiteit van de controle. Dit maakt

dat de validiteit te wensen overlaat als op basis van controles nalevingsniveaus bepaald worden.

Betrouwbaarheid
Omdat de controlerende instantie zelf kan bepalen hoeveel controles uitgevoerd worden, is het relatief eenvoudig om de betrouwbaarheid van de uitkomsten desgewenst te vergroten.

Bereik
Met name voor verplichtingen waarbij het al dan niet naleven relatief gemakkelijk te objectiveren is (woont iemand op het opgegeven adres?), kunnen controles gebruikt worden om nalevingsniveaus in kaart te brengen. Bij vormen van regelovertreding die minder gemakkelijk vast te stellen zijn, zijn controles minder bruikbaar. Dit beperkt het bereik van controles als instrument voor het meten van nalevingsniveaus.

Doorlooptijd
De doorlooptijd van controles wordt bepaald door de intensiteit van de controles, het aantal controles dat uitgevoerd wordt en het aantal controleurs dat de controles uitvoert. De doorlooptijd houdt daarmee sterk verband met andere criteria zoals betrouwbaarheid (meer controles, vergt meer doorlooptijd) en kosten (meer controleurs brengt meer kosten met zich mee, minder intensieve controles zijn goedkoper). Met voldoende inzet van capaciteit is het mogelijk om in februari van het jaar t + 1 over verslagjaar t te rapporteren.

Kosten
Uitgebreide controles (zoals huisbezoek en waarnemingen ter plaatse) zijn arbeidsintensief en brengen relatief hoge kosten met zich mee.

Draagvlak in het veld
In het veld worden aselecte controles met name gebruikt als basis om risicoprofielen uit af te leiden, of ter validatie van bestaande risicoprofielen of handhavingsmodellen. Wanneer het afleiden van risicoprofielen niet mogelijk is (vanwege lage aantallen regelovertreders), worden aselecte controles gebruikt om de mate van naleving in kaart te brengen. Met andere woorden: er is draagvlak voor het uitvoeren van aselecte controles, maar dan met name als doel om uiteindelijk de controles risicogestuurd uit te kunnen voeren.

Privacy
Het is niet zonder meer mogelijk om grote groepen cliënten intensief te controleren zonder dat er sprake is van een concreet vermoeden van fraude. Wanneer echter aannemelijk kan worden gemaakt dat de uitkeringsinstantie anders voor buitenproportionele kosten wordt gesteld, of dat er een maatschappelijk belang is voor de controles, blijken er in de praktijk mogelijkheden te zijn om aselecte controles uit te voeren.

3.6.2 Verplichtingen

In deze paragraaf concluderen we per relevante verplichting of aselecte controles een bruikbaar alternatief vormen voor zelfrapportage.

Opgave van inkomsten (eventueel van kind ouder dan 16 jaar)
Het niet opgeven van witte inkomsten kan eenvoudig gecontroleerd worden aan de hand van bestaande bestandskoppelingen. Bij controle op zwarte inkomsten leren de ervaringen van uitkeringsinstanties dat deze inkomsten niet of nauwelijks te traceren zijn. Het controleonderzoek op zich is lastig en het 'hardmaken' van de overtreding vormt eveneens een aanzienlijk probleem.

Solliciteren
Uitkeringsinstanties controleren sollicitaties doorgaans door na te gaan of het opgegeven aantal sollicitaties overeenkomt met de afspraken daarover met de cliënt. Informanten geven aan dat het niet goed mogelijk is om intensiever te controleren en tot daadwerkelijke constatering van (niet-)naleving te komen.

Trajectverplichtingen
Ook bij trajectverplichtingen geldt dat de naleving ervan moeilijk te controleren is. Uitkeringsinstanties reageren op meldingen door trajectorganisaties. Bij het uitvoeren van aselecte controles worden inspanningen op het terrein van re-integratie wel 'meegenomen', maar het vormt geen onderzoeksdoel op zich.

Verplichtingen in het kader van herstel
Deze verplichting wordt gekenmerkt door een specifiek medisch karakter. Bovendien is er sprake van een zeer subjectieve beoordeling van 'zich beter voelen', waarna dit gemeld moet worden aan de uitkeringsinstantie. Aselecte controles spitsen zich hier niet op toe.

Opgeven van gezamenlijke huishouding
De praktijkvoorbeelden die in dit hoofdstuk beschreven zijn, geven aan dat vooral om na te kunnen gaan of de feitelijke leefsituatie van een cliënt (adres, samenleefvorm) overeenkomt met de situatie zoals bekend bij de uitkeringsinstantie, aselecte controles worden uitgevoerd. Bij deze verplichting blijkt het, in tegenstelling tot de eerder beschreven verplichtingen, wel werkbaar om aan de hand van controles vast te stellen of er sprake is van naleving of niet.

3.6.3 Detection Controlled Estimation (DCE)

DCE is een methode waarmee aan de hand van controles betere schattingen van nalevingsniveaus gemaakt kunnen worden. Om deze methode toe te kunnen passen, is het nodig om een aantal kenmerken van gecontroleerden, controles en controleurs systematisch te verzamelen en te betrekken bij berekeningen van nalevingsniveaus. Op dit moment is echter onvoldoende bekend over het soort kenmerken dat daarvoor relevant is. Ook is niet bekend

of de gegevens, als ze verzameld worden voldoende aan de voorwaarden voldoen die noodzakelijk zijn voor een toepasbaar DCE-model. Daarmee concluderen wij dat het toepassen van DCE de moeite waard is als er om andere reden al voor/van controles zou worden gekozen/gebruikgemaakt. Het opzetten van controles speciaal om DCE te kunnen toepassen lijkt ons niet raadzaam vanwege de onzekerheid op succesvolle schattingsresultaten.

4 ADMINISTRATIES

4.1 Inleiding

In verband met de verstrekking van uitkeringen en met de controle op de rechtmatigheid ervan worden vele gegevens van de klant en zijn uitkering vastgelegd in uitkeringsadministraties. De gegevens worden veelal in eerste instantie uitgevraagd bij de klant en vervolgens geverifieerd, hetzij aan documenten, hetzij aan externe gegevensbestanden zoals bijvoorbeeld de GBA. De resultaten worden vastgelegd in de eigen uitkeringsadministratie. Daarnaast worden ook gegevens vastgelegd die te maken hebben met onderzoek naar regelovertreding en de uitkomsten van dit onderzoek (fraudestatistieken, debiteurenadministratie).

Denkbaar is dat deze administratieve bestanden, op zichzelf staand of in combinatie, relevant kunnen zijn voor het bepalen van het niveau van regelovertreding. In dit hoofdstuk onderzoeken we de mogelijkheden daartoe. We gaan de verschillende verplichtingen na en beschrijven de relevante gegevensbronnen naar inhoud en bruikbaarheid.

4.2 De sollicitatieverplichting (WWB, WW)

4.2.1 Inleiding

WWB- en WW- ontvangers zijn verplicht om mee te werken aan activiteiten die erop gericht zijn hun kansen op inschakeling op de reguliere arbeidsmarkt te bevorderen: de zogenaamde arbeidsverplichtingen. Simpel geformuleerd komt deze verplichting erop neer, dat de uitkeringsontvanger zich naar vermogen inspant om weer aan het werk te komen, door:

- Zelf acties te ondernemen (waaronder: solliciteren, zich beschikbaar houden voor de arbeidsmarkt, geen dingen ondernemen die arbeidsinschakeling in de weg staan).
- Mee te werken aan activiteiten die hem door de uitkeringsorganisatie worden aangeboden (trajectverplichting, zie paragraaf 1.3).

De sollicitatieverplichting is aan de orde bij de WWB en de WW. Vastgesteld moet worden of uitkeringsontvangers in voldoende mate solliciteren. Bij de WWB moeten we rekening houden met de mogelijkheid dat uitkeringsontvangers zijn vrijgesteld van deze verplichting.

4.2.2 Sollicitatieverplichting in de WWB

Gegevens over individuele bijstandsverstrekkingen worden op lokaal niveau geadministreerd. Maandelijks rapporteren gemeenten aan het CBS over

aantallen en kenmerken van hun bijstandsklanten via de Bijstandsuitkeringen statistiek (BUS). In deze statistiek is ook opgenomen of de cliënt al dan niet is vrijgesteld van deze verplichting.

Gegevens over sollicitatiegedrag van bijstandsklanten worden niet uniform vastgelegd in gemeenten, althans niet in gegevensbestanden. Dit gegeven wordt dan ook niet doorgeleverd aan het CBS.
Het niet-naleven van de sollicitatieverplichting kan via de BUS wel indirect worden benaderd, namelijk via kenmerk 27. In dat kenmerk is de reden van verlaging van de uitkering vastgelegd. Een van de antwoordmogelijkheden is: “reden op het vlak van re-integratie”.
Het overtreden van de trajectverplichting en de sollicitatieverplichting zijn in deze statistiek niet van elkaar te onderscheiden.

4.2.3 Sollicitatieverplichting en de WW

Gegevens over uitkeringsverstrekkingen WW worden vastgelegd door UWV. Op nagenoeg alle WW’ers rust de sollicitatieplicht.
Het naleven van de sollicitatieverplichting kan indirect worden benaderd via de Controle- en sanctiestatistiek Distribuerende Functie (CDF). In het zogenaamde rubrieknummer 354 wordt aangegeven om welke reden een maatregel is opgelegd. Een van de mogelijke antwoordcategorieën is: onvoldoende trachten arbeid te verkrijgen’.

4.2.4 Bespreking sollicitatieverplichting en administraties

Klanten waarvan de uitkeringsorganisatie vaststelt, dat zij de sollicitatieverplichting niet voldoende naleven, komen via BUS of CDF tot uiting. Maar voor de klanten die niet zijn onderzocht, kunnen geen directe uitspraken worden gedaan. Generalisatie moet hier uitkomst bieden.
Daarbij moeten we rekening houden met de *controlekans, detectiekans en sanctiekans*.[1]

Controlekans. Als de controles *aselect* worden uitgevoerd, bijvoorbeeld bij alle nieuwe klanten, dan is de controlekans voor alle klanten gelijk. Als deze controles *niet aselect* worden uitgevoerd, maar bijvoorbeeld op basis van risicoselectie, dan vergt het generaliseren van dergelijke resultaten gegevens over de verschillen in *controlekans*.

Detectiekans. We moeten rekening houden met eventuele verschillen in de detectiekans, wanneer we resultaten over verschillende tijdstippen met elkaar vergelijken, of tussen verschillende organisaties. Het gaat dan met name om de kwaliteit en om de intensiteit van de controles. Voor de generalisatie moeten deze gekwantificeerd kunnen worden. Dergelijke gegevens worden

[1] Controlekans: de kans dat een klant gecontroleerd wordt. Detectiekans: de kans dat een feitelijke overtreding ook uit de controle blijkt. Sanctiekans: de kans dat de vastgestelde overtreding leidt tot een sanctie of maatregel.

voor geen van beide regelingen op een uniforme manier vastgelegd. In kwalitatieve sfeer laat het zich wel beschrijven.

Sanctiekans. Voor de sanctiekans geldt eigenlijk hetzelfde als voor de detectiekans. Er moet rekening worden gehouden met eventuele verschillen naar tijd en plaats. Voor de generalisatie moeten deze gekwantificeerd kunnen worden. Dergelijke gegevens zijn niet, of alleen in meer kwalitatieve zin, voorhanden.

Kortom, deze weg is nog lang en leidt op de kortere termijn tot hoogstens een tamelijk ruwe benadering van het nalevingsniveau. Dit alles maakt het weinig aantrekkelijk om met behulp van administraties te komen tot een bepaling van de naleving van de sollicitatieverplichting.

4.3 Trajectverplichting (WWB, WW, WGA)

4.3.1 Inleiding

WWB-, WW- en WGA-gerechtigden zijn verplicht om mee te werken aan activiteiten die erop gericht zijn om hun kansen op inschakeling op de reguliere arbeidsmarkt te bevorderen: de zogenaamde arbeidsverplichtingen. Daartoe behoort ook de verplichting mee te werken aan een re-integratietraject.
We operationaliseren het begrip re-integratietraject als een vanuit de uitkeringsorganisatie (UWV, gemeente) aangeboden voorziening, gericht op de arbeidsinschakeling (of sociale activering).[2]

Bedacht moet worden dat niet alle uitkeringsontvangers zo'n re-integratietraject aangeboden krijgen. Ten eerste is er een groep van wie wordt verwacht dat zij op eigen kracht weer aan de slag komen. Ten tweede kunnen er gronden zijn om uitkeringsgerechtigden vrijstelling van een of meer van de arbeidsverplichtingen te verlenen, bijvoorbeeld vanwege persoonlijke belemmeringen of zorgtaken. Ten derde kan het zijn dat het traject in voorbereiding is: de cliënt zit op dat moment *nog* niet in een traject, maar enige tijd later vermoedelijk wel.

4.3.2 Trajectverplichting en WWB, WW

Als moet worden bepaald welk deel van de uitkeringsgerechtigden de trajectverplichting naleeft, is het van belang om zicht te hebben op de groep uitkeringsgerechtigden die daadwerkelijk een aanbod heeft gekregen. In de regel zijn de gemeenten in staat om dit in kaart te brengen. Gemeenten rapporteren hierover ook aan het CBS in de Statistiek Reïntegratie

[2] Denkbaar is om de sociale activering buiten het nalevingsonderzoek te houden, vanwege het afwijkende doel en karakter van deze verplichting.

Gemeenten (SRG).
In de SRG is geen informatie opgenomen over het niet-naleven van de trajectverplichting. Als de gemeente na controle vaststelt dat dit het geval is, en de klant daarvoor straft in de vorm van een verlaging van de uitkering, wordt dit vastgelegd in de BUS, namelijk via kenmerk 27. In dat kenmerk is vastgelegd de reden van verlaging van de uitkering. Een van de antwoordmogelijkheden is: “reden op het vlak van re-integratie”.
Net als bij de sollicitatieverplichting geldt dat het niet-naleven van de trajectverplichting dus via de BUS wel indirect kan worden benaderd.
Het overtreden van de trajectverplichting en de sollicitatieverplichting zijn in deze statistiek niet van elkaar te onderscheiden.

Het CBS meldt dat de SRG op enkele kenmerken nog niet optimaal draait. Het gaat om rapportage door gemeenten gegevens over loonkostensubsidies (een van de in te zetten voorzieningen) en de reden einde traject.

4.3.3 Trajectverplichting en WW, WGA

Analoog aan de situatie bij de WWB zijn rond de WW en WGA vergelijkbare administraties aanwezig: naast de uitkeringsadministraties voor WW en WGA is er de eerdergenoemde controle- en sanctiestatistiek (CDR) zijn er de trajectadministraties.
Daarmee zijn vergelijkbare analyses te maken als bovengenoemd, met dit verschil dat in de CDR wel onderscheid wordt gemaakt tussen overtreding van de trajectverplichting en de sollicitatieverplichting.

4.3.4 Bespreking trajectverplichting en administraties

Net als bij de sollicitatieverplichting geldt ook hier dat de te benutten administraties gegevens leveren over klanten van wie na controle is gebleken dat zij de regels hebben overtreden. Alles wat in paragraaf 2.2.4 is gezegd over *controle, detectie- en sanctiekans* is ook relevant voor het meten van de naleving van trajectverplichtingen.

4.4 Opgave inkomsten (WWB, WW, WIA, AOW, Anw)

4.4.1 Inleiding

Voor de WWB, WW, WIA en Anw geldt globaal dat inkomsten in mindering worden gebracht op de uitkering. Die inkomsten moeten dus worden vastgesteld. Voor de AKW geldt dit voor specifieke gevallen (inkomsten van kinderen 16 jaar en ouder). Voor de AOW geldt dat de partnertoeslag, die geldt in het geval dat de partner nog geen 65 jaar is, afhankelijk is van de inkomsten van deze partner.
Op uitkeringsontvangers rust de plicht dergelijke inkomsten te melden bij de desbetreffende uitkeringsorganisatie.

4.4.2 WWB

Voor controle op inkomsten van bijstandsontvangers wordt al geruime tijd gewerkt met een geautomatiseerd instrument: het Inlichtingenbureau (IB). IB koppelt nu de bijstandsklanten van gemeenten aan data van UWV, Informatiebeheergroep (IBG), Belastingdienst en de Dienst Justitiële inrichtingen van het Ministerie van Justitie (DJI). Maandelijks 'uploaden' gemeenten hun bestand en ongeveer een maand later kunnen ze het weer downloaden. Dan is het verrijkt met gegevens uit genoemde bronnen.
Langs deze weg is vast te stellen of bijstandsontvangers inkomsten uit arbeid of uit een andere uitkering hebben ontvangen, gebruikmaken van studiefinanciering, of in detentie zitten.
IB laat weten, dat op enkele uitzonderingen na, alle gemeenten zijn aangesloten bij het IB.

Gemeenten hebben langs deze weg geen zicht op inkomsten uit onderneming (zelfstandigen), of inkomsten uit arbeid waarover geen premies en/of inkomstenbelasting wordt afgedragen (zwartwerk).

Inkomsten uit zelfstandige onderneming worden geadministreerd door de Belastingdienst. Lastig is dat deze Belastingdienstgegevens pas op een laat moment worden vastgesteld.

Het zwartwerken laat zich per definitie niet benaderen via administraties.

4.4.3 WW, WIA

UWV gebruikt voor de controle op bijverdiensten de Gemeenschappelijke Verwijsindex (GVI) en gegevens van de Belastingdienst.
De GVI legt een relatie tussen de distribuerende en collecterende kant van UWV: voor iedere ontvanger van een UWV-uitkering wordt nagegaan of er sociale premies worden afgedragen, met andere woorden, of er sprake is van een dienstverband.
Met de Belastingdienst bestaan afspraken over systematische controle, min of meer vergelijkbaar aan de werkwijze bij bijstandsontvangers. Vanaf 1 januari 2006 zorgt de Belastingdienst voor de inning van de sociale premies. Daarmee zal deze controle alleen nog maar gemakkelijker worden.

4.4.4 AOW, Anw

Vastgesteld moet worden of:
- Anw-ontvangers inkomsten ontvangen, of;
- dat AOW-ontvangers verzwijgen, dat hun partner (jonger dan 65 jaar) inkomsten ontvangt.

De SVB gebruikt voor de controle op bijverdiensten gegevens van de Belastingdienst. Lastig is dat deze Belastingdienstgegevens pas op een laat

moment worden vastgesteld (jaar t+1).

Een denkbare aanvulling is bij het UWV na te gaan of er sprake is van een dienstbetrekking. Dit kan een actueler beeld opleveren van de mate van overtreding van de verplichting tot het melden van inkomsten.

4.4.5 AKW (inkomsten kinderen 16 jaar en ouder)

Vastgesteld moet worden in welke mate AKW-ontvangers verzwijgen, dat hun kinderen (16 jaar en ouder) inkomsten uit arbeid hebben. Als de inkomsten zekere grensbedragen overschrijden, vervalt in dat kwartaal het recht op kinderbijslag.
Een mogelijkheid is om uit de AKW-administratie een selectie te maken van deze kinderen, en deze aan de BD aan te leveren voor een bestandsvergelijking of om een relatie te leggen met het UWV om te bekijken of er een dienstbetrekking in het spel is.

4.4.6 Bespreking opgave inkomsten

In de afgelopen tien jaar is een grote slag gemaakt met de controle op 'witte fraude': het verzwijgen van ontvangsten door uitkeringsontvangers. Gemeenten, het UWV en de SVB krijgen 'samenloopgevallen' aangeleverd, die naar boven zijn gekomen door vergelijking van administratieve bestanden. Daarna wacht vaak nog een aanzienlijke hoeveelheid werk om van deze samenloopgevallen vast te stellen of er daadwerkelijk sprake is van verzwijgen en dus van niet-naleven van de verplichting.

De *controlekans* is binnen elk van de onderscheiden uitkeringspopulaties gelijk. Alle ontvangers van een specifieke uitkering hebben immers een gelijke kans om in onderzoek te worden meegenomen.
Voor de *detectiekans* en *sanctiekans* geldt dat niet, althans waar het gemeenten betreft. Uit onderzoek van de Inspectie Werk en Inkomen (IWI) blijkt dat er aanzienlijke verschillen bestaan tussen gemeenten als het gaat om de kwaliteit van de afhandeling van de samenloopgevallen.[3] Dat heeft onvermijdelijk gevolgen voor de mate waarin men uit de samenloopgevallen de niet-naleving detecteert en voor de mate waarin men dit bestraft.

Dat is vooral van belang voor de mogelijkheden om trends te signaleren in de regelovertreding. Als er tussen onderscheiden jaren verschillen optreden in de mate waarin en de kwaliteit waarmee gemeenten de onderzoeken ter hand nemen, dan heeft dit een zelfstandig effect op de mate van niet-naleving.

Deze beperkingen spelen bij gemeenten sterker dan bij meer gecentraliseerde organisaties als het UWV en de SVB, waar de uniformiteit van de uitvoering groter is.

[3] *Een signaal voor gemeenten; onderzoek naar de bestrijding van witte fraude*. IWI, september 2004.

4.5 Opgave gezamenlijke huishouding (WWB, AOW, Anw)

4.5.1 Inleiding

De hoogte van uitkeringen op grond van de WWB en AOW wordt gedifferentieerd naar huishoudsamenstelling. Omdat het alleenstaanden en eenoudergezinnen ontbreekt aan zekere schaalvoordelen kan hen een toeslag worden verleend. Gevolg daarvan is dat twee alleenstaanden per saldo een hogere uitkering ontvangen dan een (echt)paar. Dit kan mensen in de verleiding brengen om hun huishoudsituatie (op papier) anders voor te doen dan in werkelijkheid.
Specifiek voor de WWB geldt dat het recht op uitkering eindigt als men in detentie wordt genomen.
Specifiek voor de Anw geldt dat het recht op een uitkering eindigt op het moment dat de nabestaande hertrouwt of gaat samenwonen.

Van een gezamenlijke huishouding is sprake wanneer beide partners hun hoofdverblijf hebben op hetzelfde adres en voor elkaar zorgen, bijvoorbeeld door ieder bij te dragen aan de kosten van het huishouden (de zogenaamde 'economische verstrengeling').

4.5.2 WWB

In de reguliere uitvoeringspraktijk van de WWB wordt de opgave van het adres (hoofdverblijf) zoals die door de cliënt wordt gedaan, geverifieerd aan de GBA. Voor alle nieuwe aanvragen van de WWB wordt de informatie van de cliënt vergeleken met de administraties van DJI (vanwege detentie) en de Belasting-dienst (via het Inlichtingenbureau).
Bij nader onderzoek en opsporingsonderzoek wordt ook gebruikgemaakt van andere gegevensbanken, zoals nutsbedrijven en de RWW.

4.5.3 AOW, Anw

Ook de SVB gebruikt voor de controle op adresgegevens de databanken van de GBA, de Belastingdienst en de DJI.

4.5.4 Bespreking opgave gezamenlijke huishouding

Bij het vaststellen van de feitelijke woonsituatie wordt de klant al lang niet meer op zijn blauwe ogen geloofd; in de praktijk wordt gewerkt met een validering in de vorm van raadpleging van de GBA. Lastig punt is nu dat de kwaliteit van de GBA zelf niet onbesproken is, juist ook waar het gaat om kenmerken die voor dit onderzoek van belang zijn.

Er zijn ook andere bronnen die informatie kunnen verschaffen over de feitelijke woonsituatie. Enkele daarvan zijn inmiddels ook beproefd, maar tot nu toe vooral in de context van intensieve controles of opsporingsonderzoek

(voorbeeld: RDW, waterverbruik, woningverhuurders).

Het combineren van de adresgegevens uit verschillende administraties kan leiden tot een raming van het aantal mensen, dat op bepaalde adressen staat ingeschreven. Maar dat betekent nog niet dat er sprake is van een gezamenlijke huishouding in de zin van de wet. Daarvoor moet ook de economische verstrengeling worden aangetoond. Nader onderzoek kan dan leiden tot verdere duiding en eventueel tot het vaststellen van regelovertreding. Vanuit de opsporingspraktijk wordt benadrukt dat de economische verstrengeling zeer lastig te bewijzen is. Benutting van enkele relevante databases (bank- en verzekeringswereld) is daarbij alleen mogelijk in het kader van opsporingsonderzoek.

Een ander lastig punt van deze aanpak is het hoge 'Big Brother'-gehalte. Dat vereist een zorgvuldige aanpak, maar zelfs daarmee kent het een hoog afbreukrisico.

Dat leidt tot de conclusie dat de administratieve benadering onvoldoende aanknopingspunten biedt om tot uitspraken te komen over het verzwijgen van een gezamenlijke huishouding.

4.6 Melden herstel (WIA)

4.6.1 Melden herstel WIA

De WAO bood ondersteuning in geval van verlies van verdiencapaciteit die voortkomt uit al dan niet gedeeltelijke arbeidsongeschiktheid. Per januari 2006 is de WAO opgevolgd door de WIA. Herstel kan gevolgen hebben voor het recht op of de hoogte van de uitkering en dient dus direct te worden gemeld. Ook zijn WAO'ers verplicht mee te werken aan alles dat kan bijdragen tot het herstel. Dus moeten ze meewerken aan medische behandelingen en/of andere op herstel gerichte handelingen. Meldingen kunnen ook hierop betrekking hebben.

4.6.2 Bespreking

Klanten waarvan et UWV vaststelt, dat zij de verplichtingen niet voldoende naleven, komen via het CDF tot uiting. Maar voor de klanten die niet zijn onderzocht, kunnen geen directe uitspraken worden gedaan. Generalisatie moet hier uitkomst bieden.
Daarbij moeten we rekening houden met de *controlekans, detectiekans en sanctiekans*. Alles wat hierover is gezegd in paragraaf 4.2.4 is ook hier van toepassing.

4.7 De methodiek beoordeeld

Validiteit
Op het aspect van de validiteit zijn er de nodige knelpunten:
- Voor sommige verplichtingen hebben niet alle uitkeringsontvangers een gelijke controlekans.
- Bij vergelijking in tijd en plaats vallen problemen als gevolg van verschillen in detectiekans en sanctiekans bepaald niet uit te sluiten.

Dat maakt dat het erg lastig is om een tijdreeks van gegevens, die met deze methode worden bepaald, op een goede manier te interpreteren.
In theorie is dit op te lossen door controle-, detectie- en sanctiekans te kwantificeren. Voor de controlekans lijkt dat wel haalbaar, maar voor de detectie- en sanctiekans is het zeer twijfelachtig of dit uit te tillen is boven het niveau van een kwantificering van een kwalitatief oordeel. Dat leidt tot resultaten, die weinig precies van aard zijn en zich niet goed in termen van trends laten benoemen.

Betrouwbaarheid
In termen van herhaalbaarheid is het goed werken met bestanden. Wel een paar algemene kanttekeningen:
- De bestanden moeten waarmee wordt gewerkt, moeten naderhand niet worden gewijzigd, of deze wijzigingen moeten traceerbaar en gedateerd zijn aangebracht.
- Een behoorlijk deel van de relevante statistieken en administratieve bestanden zijn in de afgelopen paar jaar gewijzigd, niet allemaal zijn ze inmiddels over hun volle breedte als stabiel te bestempelen (bijvoorbeeld: SRG).
- Er is twijfel over de kwaliteit van de GBA, juist ook op kenmerken die voor deze studie relevant zijn. Belangrijkste knelpunt is dat opgave van de burger niet altijd geverifieerd wordt of kan worden (voorbeeld: juistheid van opgegeven woonadres).

Bereik
Het bereik van deze methode is wisselend voor het bepalen van de naleving van de onderscheiden verplichtingen.
De ene verplichting (bijvoorbeeld: opgave van inkomsten) is veel beter gedekt door administraties dan de ander (bijvoorbeeld: sollicitatieverplichting).

Doorlooptijd
Gegevens over verslagjaar t dienen beschikbaar te zijn in februari van het jaar t + 1.
Het verwerken van gegevens in een administratie neemt natuurlijk zijn tijd. De vulling loopt daarom per definitie achter bij de werkelijkheid. In geval van statistieken vindt er meestal ook nog overleg plaats tussen de leverende en verzamelende partij. Daarmee kan het enkele weken tot vele maanden duren

voordat een statistiek op orde is en vrijgegeven wordt. Als gebruik wordt gemaakt van administraties die het kalenderjaar volgen, dan is rapportage in februari van jaar t + 1 zeer onwaarschijnlijk.

Draagvlak in het veld
Het praktische, arbeidsextensieve karakter van de aanpak spreekt mensen aan. Anderen twijfelen vanwege beperkingen in sfeer van de validiteit en betrouwbaarheid.

Privacybescherming
Dit aspect is verkend maar niet grondig onderzocht. Daaruit zijn ons op het vlak van de privacybescherming geen onoverkomelijke problemen gebleken. Wel dient er het nodige te worden geregeld voor de gegevens gebruikt kunnen worden, waarmee ook de nodige tijd gemoeid kan zijn. Verder dienen zich specifieke voorwaarden aan, waaronder gegevens kunnen worden gebruikt. Bijvoorbeeld voor het CBS geldt dat er geen op naam gestelde (tot het individu herleidbare) gegevens het gebouw mogen verlaten. Dat kan belangrijke gevolgen hebben voor de inrichting van het onderzoek (‘wie doet wat en waar?’).

Kosten
De kosten zijn relatief laag, omdat er geen (arbeidsintensieve) primaire dataverzameling nodig is.

5 METHODEN EN VERPLICHTINGEN

5.1 Inleiding

In dit hoofdstuk werken we de specifieke kenmerken van de drie methodes uit voor de onderscheiden verplichtingen. Algemene beschrijvingen van de prestaties van de onderscheiden methodes zijn gegeven in eerdere hoofdstukken:
- Paragraaf 2.4 over zelfrapportage.
- Paragraaf 3.6 over aselecte controles.
- Paragraaf 4.7 over de administratieve benadering.

5.2 Verplichting tot solliciteren

5.2.1 Inleiding

Relevante regelingen: WW, WWB.

Gewenste indicatoren jaarverslag SZW:
- Nalevingsverplichting sollicitatieverplichtingen.

Huidige indicatoren POROSZ, gebruikt in SZW-begroting 2005:
- Heeft u in de afgelopen twaalf maanden wel eens langere tijd bewust weinig gesolliciteerd?

Nagegaan moet worden in welke mate uitkeringsontvangers zich houden aan de verplichting tot solliciteren. Er moet rekening worden gehouden met het feit dat een deel van de WWB-ontvangers zijn vrijgesteld van deze verplichting.

5.2.2 Bespreking van de methodes

Zelfrapportage

Het naleven van de verplichting tot solliciteren kan zonder veel omhaal aan personen worden gevraagd, zo blijkt al uit de POROSZ-ervaring. Hoewel empirisch niet bewezen, lijkt dit ook geen erg gevoelige vraag te zijn. In dat geval is de directe bevraging hier een serieus alternatief voor de randomized-responsemethode.
Wel is het raadzaam te toetsen of de veronderstelling over de gevoeligheid klopt. Dat kan door dit onderwerp eenmalig uit te vragen met de directe bevraging en de randomized-responsemethode.

Een lastig punt is de interpretatie van de vraag of men bewust weinig heeft gesolliciteerd. Omdat niet helder genormeerd is wat in dit verband verstaan moet worden onder ‘weinig’, zal de respondent zich baseren op eigen opvattingen daarover.

In geval van de bijstandsontvangers kan worden overwogen deze vraag alleen voor te leggen aan diegenen die de verplichting ook daadwerkelijk kunnen overtreden. Dat kan met een eenvoudige selectievraag (“Bent u door de gemeente vrijgesteld van de sollicitatieplicht?”).
Het werken met een selectievraag is echter niet mogelijk in combinatie met randomized response.
Om het werken met een selectievraag te vermijden, kan worden overwogen om gegevens van gemeenten of CBS (de BUS) te gebruiken als steekproefkader. Echter, gebruik van gemeentedata levert problemen op met de generalisering, en CBS-data hebben beperkingen in de zin dat tot individuele personen herleidbare gegevens het CBS-gebouw niet mogen verlaten.

Validiteit uitkomsten
Op grond van het bovenstaande en ervan uitgaande dat het sollicitatiegedrag een relatief minder gevoelig onderwerp is, beoordelen wij de validiteit als redelijk, voor zowel de randomized-responsemethode als de directe bevraging. Voorwaarde is dat gebruik wordt gemaakt van adequate steekproefkaders.

Betrouwbaarheid uitkomsten
Steekproefonderzoek gaat gepaard met betrouwbaarheidsmarges. Die marges zullen, gegeven de steekproefomvang, bij de directe bevraging wat kleiner zijn dan bij de randomized-responsemethode. Wij beoordelen de betrouwbaarheid als respectievelijk redelijk en matig.

Doorlooptijd
Gegevens over verslagjaar t kunnen beschikbaar zijn in februari van het jaar t + 1.

Kosten
Deze methode is relatief duur, in ieder geval de randomized-responsevariant. Mogelijk is echter ook de directe bevraging een kostbare aangelegenheid. Vanwege de beperktere waarneming van regelovertreding zullen ook de veranderingen daarin kleiner zijn. Om dat met directe bevraging te kunnen waarnemen, zijn weer grotere steekproefaantallen nodig. Welk van de twee varianten duurder uitvalt, is op voorhand niet aan te geven.

Draagvlak in het veld
Geen specifieke opmerkingen. Hetgeen betekent: er is voldoende draagvlak.

Privacybescherming
Geen specifieke opmerkingen. Dat betekent dat de benadering via access pools geen privacyproblemen oplevert. Als gebruik wordt gemaakt van databanken van de Belastingdienst of het UWV zijn er geen onoverkomelijke problemen.

Aselecte controles

Uitkeringsinstanties controleren sollicitaties doorgaans door na te gaan of het opgegeven aantal sollicitaties overeenkomt met de afspraken daarover met de cliënt. De kwaliteit van de sollicitaties laat zich lastig vaststellen, omdat moeilijk te bepalen is wat kan gelden als 'voldoende' solliciteren. Heldere normen daarvoor ontbreken.

Informanten geven aan dat meer intensieve, aselecte controles niet vaak leiden tot (extra) waarnemingen van niet-naleving.
Dat heeft onder meer te maken met het feit dat deze controlesoort gebonden is aan het 'hier en nu', terwijl solliciteren niet iets is waarmee men dag-in-dag-uit actief is. Veel van de regelovertreding valt dus buiten de waarneming van de controle, hetgeen leidt tot lage prevalentiecijfers.
Daarnaast is deze methode sterk afhankelijk van de beoordeling door individuele controleurs. Gegeven de aanzienlijke interpretatieruimte rond de sollicitatieverplichting kan dit leiden tot grote individuele verschillen in de beoordeling en daarmee in waarneming van regelovertreding.

Validiteit
Op grond van het bovenstaande beoordelen we de validiteit van aselecte controles op sollicitatiegedrag als matig. In een omgeving waarin de controles in sterke mate centraal worden aangestuurd en geprotocolleerd, is de validiteit wellicht nog wat beter.

Betrouwbaarheid
Omdat de controlerende instantie zelf kan bepalen hoeveel controles uitgevoerd worden, is het relatief eenvoudig om de betrouwbaarheid van de uitkomsten te beïnvloeden. Om met enige precisie uitspraken te kunnen doen, zal een aanzienlijke groep moeten worden gecontroleerd.
Bij de generalisatie van de steekproefresultaten naar populatieniveau moeten de nodige aannames worden gehanteerd, hetgeen ten koste gaat van de precisie van de onderzoeksresultaten. Veelal zal de generalisatie bestaan uit twee stappen, namelijk van de onderzoeksgroep naar de gemeente of vestiging en vervolgens van dat niveau naar het populatieniveau.
Wij beoordelen de betrouwbaarheid als matig.

Kosten
Het gaat bij aselecte controles om tamelijk arbeidsintensieve werkzaamheden. Dat zal dan ook gepaard gaan met aanzienlijke kosten. Dit punt geldt overigens niet specifiek voor deze verplichtingen, maar voor alle verplichtingen.

Doorlooptijd
Geen bijzonderheden ten opzichte van aselecte controles bij andere nalevingsverplichtingen.

Draagvlak in het veld
Informanten geven aan dat het niet goed mogelijk is om bij controles op het nakomen van de verplichting tot solliciteren, intensiever te werk te gaan dan men bij reguliere controles al doet. Het draagvlak beoordelen wij als matig.

Privacy
Geen specifieke opmerkingen. Het is niet zonder meer mogelijk om groepen cliënten intensief te controleren zonder dat er sprake is van een concreet vermoeden van fraude. Wanneer echter aannemelijk kan worden gemaakt dat de uitkeringsinstantie anders voor buitenproportionele kosten wordt gesteld, of er een maatschappelijk belang is voor de controles, blijken er in de praktijk mogelijkheden te zijn om aselecte controles uit te voeren.

Administraties *Validiteit*
Klanten waarvan de uitkeringsorganisatie vaststelt dat zij de sollicitatie-verplichting niet voldoende naleven, komen via bijstandsuitkeringenstatistiek BUS (of daaraan ten grondslag liggende lokale administraties) of fraude-statistiek UWV (CDF) tot uiting. Voor de overige klanten kan echter niet worden gesteld dat zij de regels wel naleven. Bij generalisatie moeten we dan rekening houden met de *controlekans, detectiekans en sanctiekans.* Het probleem is dat deze kansen zich niet precies laten kwantificeren en ook niet constant zijn in de loop der tijd. Derhalve leidt deze methode hoogstens tot een ruwe benadering van het nalevingsniveau.
De methode levert daarom weinig valide resultaten.

Betrouwbaarheid
Bij het gebruik van de administraties worden in de regel geen steekproeven gebruikt, generaliseren en betrouwbaarheidsmarges zijn niet aan de orde. Wij beoordelen wij de betrouwbaarheid als goed.

Kosten
Geen specifieke opmerkingen, maar in het algemeen geldt dit als relatief gunstig, vanwege het feit dat gebruik wordt gemaakt van reeds beschikbare informatie.

Draagvlak in het veld
In het veld leven geen hoge verwachtingen van het resultaat van deze methode.

Doorlooptijd
Met het beschikbaar komen van bestanden is in het algemeen de nodige tijd gemoeid. Als gebruik moet worden gemaakt van een jaarbestand over jaar t, dan is onwaarschijnlijk dat in februari van jaar t+1 gerapporteerd kan worden.

Privacybescherming
Geen onoplosbare problemen.

5.2.3 Conclusie

Voor het bepalen van de naleving van de sollicitatieverplichting, levert zelf-rapportage meer valide resultaten dan beide andere methoden. Juist omdat het niet zo'n gevoelig item lijkt te betreffen, is de verwachte meerwaarde van de randomized-responsemethode beperkt ten opzichte van de directe bevraging.
Het draagvlak in het veld is ook groter voor de zelfrapportage. Een lastig punt blijft het ontbreken van een heldere normering voor het beoordelen van sollicitatiegedrag, hetgeen zich bij alle beschreven methodes manifesteert. Qua kosten scoort de administratieve benadering het best, maar in doorlooptijd en validiteit kent het serieuze problemen.

Het lijkt plausibel te veronderstellen dat de sollicitatieverplichting weinig gevoelig ligt bij respondenten. Of dat werkelijk zo is, moet blijken uit empirisch onderzoek. Dat kan door dit onderwerp eenmalig zowel via directe bevraging als via de randomized-responsemethode te onderzoeken.

Om tot resultaten te komen met een behoorlijke validiteit moet gebruik worden gemaakt van adequate steekproefkaders.

Daarnaast kunnen de resultaten uit de BUS en CDF worden gebruikt als indicatie voor de ondergrens van de prevalentie.

5.3 Nakomen trajectafspraken

5.3.1 Inleiding

Relevante regelingen: WW, WWB.

Gewenste indicatoren jaarverslag SZW:
- Nalevingsniveau trajectverplichting.

Huidige indicatoren POROSZ, gebruikt in SZW-begroting 2005:
- Trajectafspraken: geen indicator.

Nagegaan moet worden in welke mate uitkeringsontvangers zich houden aan afspraken, die met hen zijn gemaakt over het volgen van een reïntegratietraject. Daarbij moet rekening worden gehouden met het feit dat een (aanzienlijk) deel van de uitkeringsontvangers geen traject volgt.

5.3.2 Bespreking van de methodes

Zelfrapportage

Validiteit uitkomsten
Er is in POROSZ nog geen ervaring opgedaan met het uitvragen van deze nalevingsverplichting.
De trajectverplichting lijkt veel gemeen te hebben met de sollicitatie-verplichting. Het lijkt geen al te gevoelig item te zijn. Veel van wat over de sollicitatieverplichting is gezegd, kan worden betrokken op de traject-verplichting. Dat betekent dat ook bij de bevraging van dit item de directe bevraging zich aandient als alternatief voor de randomized-responsemethode. Het betekent ook dat het lastig kan zijn om de naleving van deze verplichting helder te normeren.

Anders dan bij de sollicitatieplicht is bij de trajectverplichting het selecteren van de doelgroep lastiger. De reden daarvoor is dat de verplichting zich niet makkelijk laat vertalen in een selectievraag die zowel begrijpelijk als doel-treffend is. Aan dit probleem kan worden tegemoetgekomen door de re-integratieadministraties van het UWV en gemeenten (al dan niet via het CBS) te gebruiken als steekproefkader.
Uitgaande van een dergelijk steekproefkader beoordelen wij de validiteit van beide varianten als redelijk.

Betrouwbaarheid uitkomsten
Als bij de sollicitatieverplichting, dat wil zeggen matig (randomized-responsemethode) tot redelijk (directe bevraging), gegeven een zekere steekproefomvang.

Doorlooptijd
Als bij de sollicitatieverplichting, dat wil zeggen relatief gunstig.

Kosten
Als bij de sollicitatieverplichting, dat wil zeggen relatief kostbaar.

Draagvlak in het veld
Als bij de sollicitatieverplichting, dat wil zeggen redelijk.

Privacybescherming
Als bij de sollicitatieverplichting, dat wil zeggen zonder onoverkomelijke problemen.

Aselecte controles

Validiteit
Ook bij trajectverplichtingen geldt dat de naleving ervan moeilijk te controleren is. Uitkeringsinstanties reageren op meldingen door trajectorganisaties en voeren zelf geen aselecte controles uit naar deze verplichting. Wanneer dit wel zou gebeuren, zal het niet in alle gevallen gemakkelijk zijn om cliënten op een specifiek moment in tijd en ruimte te lokaliseren (uitkeringsorganisatie staat op afstand; meerdere re-integratiebedrijven betrokken, IRO). Bovendien is het

niet eenvoudig om vast te stellen of het eventueel niet aanwezig zijn de cliënt te *verwijten* valt. Het is immers mogelijk dat er in de informatievoorziening aan de cliënt fouten zijn gemaakt, zodat de cliënt niet weet waar en wanneer hij aanwezig moet zijn.
Het is dan ook lastig om tot een uniforme beoordeling te komen. De controles zullen immers worden uitgevoerd door een aanzienlijk aantal controleurs en de beoordelingscriteria zullen zich niet altijd volledig laten objectiveren en expliciteren.
We beoordelen de validiteit van aselecte controles op sollicitatiegedrag als matig. In een omgeving waarin de controles in sterke mate centraal worden aangestuurd en geprotocolleerd, is de validiteit wellicht nog wat beter.

Betrouwbaarheid
Als bij de sollicitatieverplichting, dat wil zeggen: wij beoordelen de betrouwbaarheid als matig.

Doorlooptijd
Geen bijzonderheden ten opzichte van aselecte controles bij andere nalevingsverplichtingen.

Kosten
Als bij de sollicitatieverplichting, dat wil zeggen: uitgebreide controles brengen relatief hoge kosten met zich mee.

Draagvlak in het veld
Uitkeringsinstanties reageren op meldingen door trajectorganisaties en voeren zelf geen aselecte controles uit naar deze verplichting. Dit is in verband met de moeite die het kost om niet-naleving vast te stellen en vervolgens te bepalen wie verwijtbaar is (cliënt, uitkeringsinstantie of trajectorganisatie).
We beoordelen het draagvlak als matig.

Privacy
Geen bijzonderheden ten opzichte van aselecte controles bij andere nalevingsverplichtingen.

Administraties

Validiteit
Met welke cliënten een trajectafspraken zijn gemaakt, valt af te leiden uit reïntegratieadministraties van gemeenten en UWV. Of cliënten zich aan die verplichting houden blijkt daaruit echter niet. Alleen wanneer bij cliënten wordt vastgesteld dat zij de trajectverplichting niet voldoende naleven, komt dit via administraties (BUS of daaraan ten grondslag liggende lokale administraties; CDF) tot uiting. Voor de overige klanten kan niet worden gesteld dat zij de regels wel naleven.

Bij generalisatie moeten we dan rekening houden met de *controlekans, detectiekans en sanctiekans.* Probleem is, dat deze kansen zich niet precies laten kwantificeren en ook niet constant zijn in de loop der tijd. Derhalve leidt deze methode hoogstens tot een ruwe benadering van het nalevingsniveau. De methode levert daarom weinig valide resultaten.

Betrouwbaarheid
Bij het gebruik van de administraties worden in de regel geen steekproeven gebruikt, generaliseren en betrouwbaarheidsmarges zijn niet aan de orde. Wij beoordelen daarom de betrouwbaarheid als goed.

Doorlooptijd
Met het beschikbaar komen van bestanden is in het algemeen de nodige tijd gemoeid. Als gebruik moet worden gemaakt van een jaarbestand over jaar t, dan is onwaarschijnlijk dat in februari van jaar t+1 gerapporteerd kan worden.

Kosten
Als bij de sollicitatieverplichting, dat wil zeggen relatief gunstig.

Draagvlak in het veld
Als bij de sollicitatieverplichting, dat wil zeggen dat men er geen hoge verwachtingen van heeft.

Privacybescherming
Als bij de sollicitatieverplichting, dat wil zeggen dat zich geen onoplosbare problemen hebben aangediend.

5.3.3 Conclusie

Als bij de sollicitatieverplichting, met dit verschil dat het selecteren van de doelgroep voor de trajectverplichting met een selectievraag een lastige zaak is.
Om die reden kan worden overwogen om re-integratieadministraties te gebruiken als (aanvullend) steekproefkader.

5.4 Opgave van inkomsten uit arbeid

5.4.1 Inleiding

Relevante regelingen WWB, WW, WIA, AOW, Anw.

Gewenste indicatoren jaarverslag SZW:
- nalevingsniveau opgave inkomsten uit arbeid.

Huidige indicatoren POROSZ:
- verzwijgen van inkomsten uit kleinere klusjes voor of via bekenden;
- verzwijgen van inkomsten uit zwartwerk;
- verzwijgen van inkomsten uit baan of uitzendwerk (wit).

5.4.2 Bespreking van de methodes

Zelfrapportage

Validiteit
Het lijkt ons plausibel te veronderstellen dat vragen naar het verzwijgen van inkomsten gevoeliger liggen dan de vragen naar sollicitatie- en trajectverplichting. Dat betekent een grotere meerwaarde van de randomized-responsevariant ten opzichte van de directe bevraging. Het onderscheid tussen klusjes, wit- en zwartwerk lijkt ons voor klanten moeilijk te maken, hetgeen ten koste gaat van de validiteit.
De validiteit van de directe bevraging beoordelen we als matig, de randomized-responsemethode als redelijk.

Betrouwbaarheid uitkomsten
Als bij de sollicitatieverplichting, dat wil zeggen matig (randomized-responsemethode) tot redelijk (directe bevraging), gegeven een zekere steekproefomvang.

Doorlooptijd
Als bij de sollicitatieverplichting, dat wil zeggen relatief gunstig.

Kosten
Als bij de sollicitatieverplichting, dat wil zeggen relatief kostbaar.

Draagvlak in het veld
Als bij de sollicitatieverplichting, dat wil zeggen redelijk.

Privacybescherming
Als bij de sollicitatieverplichting, dat wil zeggen zonder onoverkomelijke problemen.

Aselecte controles

Validiteit
Het niet opgeven van witte inkomsten kan eenvoudig gecontroleerd worden aan de hand van bestaande bestandskoppelingen. Daar voegen aselecte controles niets aan toe.
Bij controle op zwarte inkomsten leren de ervaringen van uitkeringsinstanties dat deze inkomsten niet of nauwelijks te traceren zijn. Het controleonderzoek op zich is lastig en het ‘hardmaken’ van de overtreding vormt eveneens een aanzienlijk probleem. Er worden wel successen gemeld met waarnemingen ter plaatse, die plaatsvinden in de context van de gezamenlijke onderzoeken die worden gedaan in de interventieteams (voormalige RIF’s) of van onderzoeken van de Arbeidsinspectie. Alleen zeer intensieve controles kunnen in sommige gevallen uitkomst bieden.

Veel regelovertreding blijft derhalve onopgemerkt. Ook zijn er nauwelijks houdbare handvatten voor het generaliseren van de resultaten naar populatieniveau. Dit maakt dat wij de validiteit beoordelen als matig.

Betrouwbaarheid
Zoals gezegd: er zijn nauwelijks houdbare handvatten voor het generaliseren van de resultaten naar populatieniveau. Dit maakt dat wij de betrouwbaarheid beoordelen als matig.

Doorlooptijd
De doorlooptijd van controles wordt bepaald door de intensiteit van de controles, het aantal controles dat uitgevoerd wordt en het aantal controleurs dat de controles uitvoert. De doorlooptijd houdt daarmee sterk verband met andere criteria zoals betrouwbaarheid (meer controles, vergt meer doorlooptijd) en kosten (meer controleurs brengen meer kosten met zich mee, minder intensieve controles zijn goedkoper).
Met voldoende inzet van capaciteit is het mogelijk om in februari van het jaar t + 1 over verslagjaar t te rapporteren.

Kosten
Uitgebreide controles (zoals huisbezoek en waarnemingen ter plaatse) zijn arbeidsintensief en brengen relatief hoge kosten met zich mee. Om zwartwerk aan te kunnen tonen, zijn intensieve en dus dure controles nodig.

Draagvlak in het veld
De meeste respondenten in het veld geven aan dat zwartwerk een belangrijk handhavingsspeerpunt is. Men is echter van mening dat aselecte controles weinig houvast bieden om het niveau van regelovertreding op dit punt vast te stellen. Dit vanwege de moeite om bij een controle objectief vast te stellen of er sprake is van zwarte inkomsten.

Privacy
Geen bijzonderheden ten opzichte van aselecte controles bij andere nalevingsverplichtingen.

Administraties

In de afgelopen tien jaar is een grote slag gemaakt met de controle op ‘witte fraude’: het verzwijgen van ontvangsten door uitkeringsontvangers. Gemeenten, UWV en SVB krijgen ‘samenloopgevallen’ aangeleverd, die naar boven zijn gekomen door vergelijking van administratieve bestanden. Daarna wacht vaak nog een aanzienlijke hoeveelheid werk om van deze samenloopgevallen vast te stellen of er daadwerkelijk sprake is van verzwijgen en dus van niet-naleven van de verplichting. Over witte fraude wordt jaarlijks gerapporteerd in bijstandsfraudestatistiek (BFS) en CDF (zie verder onder Administraties).
De *controlekans* is binnen elk van de onderscheiden uitkeringspopulaties gelijk. Alle ontvangers van een specifieke uitkering hebben immers een gelijke kans om in onderzoek te worden meegenomen.

Voor de *detectiekans* en *sanctiekans* geldt dat niet, althans waar het gemeenten betreft. Uit onderzoek van de Inspectie Werk en Inkomen (IWI) blijkt dat er aanzienlijke verschillen bestaan tussen gemeenten als het gaat om de kwaliteit van de afhandeling van de samenloopgevallen.[1] Dat heeft onvermijdelijk gevolgen voor de mate waarin men uit de samenloopgevallen de niet-naleving detecteert en voor de mate waarin men dit bestraft.

Dat is vooral van belang voor de mogelijkheden om trends te signaleren in de regelovertreding. Als er tussen onderscheiden jaren verschillen optreden in de mate waarin en kwaliteit waarmee gemeenten de onderzoeken ter hand worden genomen, dan heeft dit een zelfstandig effect op de mate van niet-naleving.

Soortgelijke kanttekeningen kunnen ook worden gemaakt voor het UWV en de SVB, zij dat daar de uniformiteit van de werkwijze minder risico's loopt.

Validiteit en betrouwbaarheid
Op grond van het bovenstaande beoordelen we zowel de validiteit als de betrouwbaarheid van de methode voor het meten van verzwijgen van zwarte inkomsten als slecht.
Voor het verzwijgen van witte inkomsten zijn zowel de validiteit als de betrouwbaarheid goed.

Doorlooptijd
Geen specifieke opmerkingen.

Kosten
Geen specifieke opmerkingen.

Draagvlak in het veld
Geen specifieke opmerkingen.

Privacybescherming
Er wordt al lange tijd discussie gevoerd over uitbreiding van het aantal gegevensbanken dat gebruikt kan worden in het kader van controles door uitkeringsinstellingen. Daarin is meer helderheid gekomen met de brief van de Staatssecretaris van SZW van 23 december 2005, waarin extra mogelijkheden worden gedefinieerd (bankrekeningnummers via de Belastingdienst, ruimere raadpleging van de RDW).

5.4.3 Conclusie

Als het gaat om het waarnemen van het niet-melden van zwarte inkomsten, heeft de administratieve benadering per definitie niets te bieden. De aselecte

[1] *Een signaal voor gemeenten; onderzoek naar de bestrijding van witte fraude*. IWI, september 2004.

controles worden gekenmerkt door een zeer beperkte validiteit. Het generaliseren van de resultaten is problematisch en betrouwbaarheid beperkt. Eigenlijk lijkt alleen de methode van zelfrapportage in staat om een redelijk valide en betrouwbaar beeld van de prevalentie van het verzwijgen van zwarte inkomsten te leveren. Omdat het gaat om een relatief gevoelig gegeven, is aannemelijk dat de randomized-responsevariant meerwaarde biedt boven directe bevraging.

Daarnaast kunnen de resultaten uit de Fraudestatistiek WWB en CDF worden gebruikt als indicatie voor de ondergrens van de prevalentie van de witte fraude. Gezien het lastig te interpreteren onderscheid in POROSZ tussen klusjes, witte en zwarte inkomsten kunnen zich daarbij wel problemen voordoen met de vergelijkbaarheid.

5.5 Verplichting opgeven gezamenlijke huishouding

5.5.1 Inleiding

Relevante regelingen AOW, Anw.

Gewenste indicatoren jaarverslag SZW:
- Melden van voren van een gezamenlijke huishouding.

Huidige indicatoren POROSZ:
- Heeft u ooit onjuiste informatie of onvolledige informatie gegeven over woonsituatie, gezinssamenstelling, burgerlijke staat?
- Samen leven met partner zonder doorgeven aan SVB?

5.5.2 Bespreking van de methodes

Zelfrapportage *Validiteit*

De gevoeligheid van dit onderwerp lijkt relatief groot, vergelijkbaar met het niet opgeven van inkomsten. Om die reden is het aannemelijk dat de randomized-responsevariant meerwaarde biedt boven directe bevraging.
Een lastig punt is de interpretatie van de vragen die hier aan de orde zijn. Wat moet in dit verband bijvoorbeeld worden verstaan onder ‘samenleven met een partner’? Vanwege het ontbreken van helderheid zal de respondent zich baseren op eigen opvattingen daarover.
Een eventuele selectievraag is hier niet zonder problemen te formuleren, maar een grove selectie is wel te maken op basis van de formele inschrijvingspositie in de GBA als alleenstaande.
De validiteit van de directe bevraging beoordelen we als matig, de randomized-responsemethode als redelijk.

Betrouwbaarheid uitkomsten
Als bij de sollicitatieverplichting, dat wil zeggen matig (randomized-responsemethode) tot redelijk (directe bevraging), gegeven een zekere steekproefomvang.

Doorlooptijd
Als bij de sollicitatieverplichting, dat wil zeggen matig.

Kosten
Als bij de sollicitatieverplichting, dat wil zeggen relatief kostbaar.

Draagvlak in het veld
Als bij de sollicitatieverplichting, dat wil zeggen redelijk.

Privacybescherming
Als bij de sollicitatieverplichting, dat wil zeggen zonder onoverkomelijke problemen.

Aselecte controles

Validiteit
In de praktijk worden aselecte controles vooral uitgevoerd om na te kunnen gaan of de feitelijke leefsituatie van een cliënt (adres, samenleefvorm) overeenkomt met de situatie zoals bekend bij de uitkeringsinstantie. Bij deze verplichting blijkt het, in tegenstelling tot de eerder beschreven verplichtingen, wel werkbaar om aan de hand van controles vast te stellen of er sprake is van naleving of niet. Dat heeft veel te maken met het specifieke eigenschappen van deze verplichting. Immers, solliciteren is veel minder dan samenwonen een bezigheid die op elk moment is waar te nemen.
Qua validiteit scoort deze methode hier in ieder geval beter dan bij de eerder beschreven verplichtingen. Een lastig punt blijft de afhankelijkheid van de kwaliteit en uniformiteit van controleurs en controles.
Op grond van het bovenstaande beoordelen we de validiteit van aselecte controles op sollicitatiegedrag als matig. In een omgeving waarin de controles in sterke mate centraal worden aangestuurd en geprotocolleerd is de validiteit wellicht nog wat beter.

Betrouwbaarheid
Geen bijzonderheden ten opzichte van aselecte controles bij andere nalevingsverplichtingen. Wij beoordelen de betrouwbaarheid als matig.

Doorlooptijd
Geen bijzonderheden ten opzichte van aselecte controles bij andere nalevingsverplichtingen.

Kosten
Uitgebreide controles (zoals huisbezoek en waarnemingen ter plaatse) zijn arbeidsintensief en brengen relatief hoge kosten met zich mee.

Draagvlak in het veld
In het veld worden aselecte controles met name gebruikt als basis om risicoprofielen uit af te leiden, of ter validatie van bestaande risicoprofielen of handhavingsmodellen. Wanneer het afleiden van risicoprofielen niet mogelijk is (vanwege lage aantallen regelovertreders), worden aselecte controles gebruikt om de mate van naleving in kaart te brengen. In de meeste praktijkvoorbeelden die wij vonden van aselecte controles, betrof het controles naar het adres en de leefsituatie van de cliënt.
Met andere woorden: er is draagvlak voor het uitvoeren van aselecte controles naar woon- en leefsituatie, maar het gaat dan met name om controles die tot doel hebben om uiteindelijk de controles risicogestuurd uit te kunnen voeren.

Privacy
Geen bijzonderheden ten opzichte van aselecte controles bij andere nalevingsverplichtingen.

Administraties

Bij het vaststellen van de feitelijke woonsituatie wordt de klant al lang niet meer op zijn blauwe ogen geloofd. In de praktijk wordt gewerkt met een validering in de vorm van raadpleging van de GBA. Een lastig punt is nu dat de kwaliteit van de GBA zelf niet onbesproken is, juist ook waar het gaat om kenmerken die voor dit onderzoek van belang zijn.

Er zijn zeker ook andere bronnen die informatie kunnen verschaffen over feitelijke woonsituatie. Enkele daarvan zijn inmiddels ook beproefd, maar tot nu toe vooral in de context van intensieve controles of opsporingsonderzoek (voorbeeld: RDW, waterverbruik, woningverhuurders).

Het combineren van de adresgegevens uit verschillende administraties kan leiden tot een raming van het aantal mensen dat op bepaalde adressen staat ingeschreven. Dat betekent echter nog niet dat er sprake is van gezamenlijke huishouding in de zin van de wet. Daarvoor moet ook de economische verstrengeling worden aangetoond. Nader onderzoek kan dan leiden tot verdere duiding en eventueel tot het vaststellen van regelovertreding. Vanuit de opsporingspraktijk wordt benadrukt dat de economische verstrengeling zeer lastig te bewijzen is. Benutting van enkele relevante databases (bank- en verzekeringswereld) is daarbij alleen mogelijk in het kader van opsporingsonderzoek.

Een ander lastig punt van deze aanpak is het hoge ‘Big Brother’-gehalte. Dat vereist een zorgvuldige aanpak, maar zelfs daarmee kent het een hoog afbreukrisico.

Dat leidt tot de conclusie dat de administratieve benadering onvoldoende aanknopingspunten biedt om tot uitspraken te komen over het verzwijgen van gezamenlijke huishouding.

Validiteit en betrouwbaarheid
Op grond van het bovenstaande beoordelen we de validiteit en betrouwbaarheid als beperkt.

Doorlooptijd
Geen specifieke opmerkingen.

Kosten
Geen specifieke opmerkingen.

Draagvlak in het veld
Als bij sollicitatieverplichting.

Privacybescherming
Als onder 'verzwijgen van inkomsten'.

5.5.3 Conclusie

Bij het waarnemen van het niet-opgeven van samenwonen zijn goede ervaringen opgedaan met *aselecte controles*, zo leren ervaringen van de SVB. Een risico blijft de afhankelijkheid van de kwaliteit en uniformiteit van controleurs en controles. Dit risico is groter naarmate de handelingsruimte (discretionaire bevoegdheden) van controleurs in een organisatie groter is. Waar dit risico beperkt is, kan deze methode redelijk valide resultaten opleveren. De methode is wel kostbaar.
Dat laatste geldt tevens voor de *zelfrapportage,* die ook op dit onderwerp in staat is om redelijk valide en betrouwbare resultaten op te leveren. Gezien de gevoeligheid van het onderwerp zal de randomized-responsevariant meerwaarde bieden boven de directe bevraging, in termen van validiteit.
De *administratieve benadering* biedt op dit moment onvoldoende aanknopingspunten om tot uitspraken te komen over het verzwijgen van gezamenlijke huishouding.

5.6 Melden herstel, meewerken aan herstelactiviteiten

5.6.1 Inleiding

Relevante regelingen WIA, Wajong, WAZ.

Gewenste indicatoren jaarverslag SZW:
- Nalevingsverplichting in het kader van herstel.

Huidige indicatoren POROSZ:
- medische diagnose afname arbeidsongeschiktheid niet doorgegeven;
- zieker voorgedaan tijdens controles;
- bij werk bemerkt herstel van arbeidsongeschiktheid niet doorgegeven;
- in staat meer te werken zonder dit door te geven.

5.6.2 Bespreking van de methodes

Zelfrapportage

Validiteit
Gevoeligheid van dit onderwerp lijkt relatief beperkt, vergelijkbaar met de sollicitatieverplichtingen. Op dit onderwerp is het lastig om begrijpelijke en doeltreffende vragen te stellen. Immers, wat moet worden verstaan onder herstel? Hoe lang moet dat aanhouden, hoeveel vooruitgang is relevant? De validiteit lijkt dus beperkt.

Ook een eventuele selectievraag laat zich hier lastig formuleren.

Betrouwbaarheid uitkomsten
Als bij de sollicitatieverplichting, dat wil zeggen relatief gunstig.

Doorlooptijd
Als bij de sollicitatieverplichting, dat wil zeggen matig.

Kosten
Als bij de sollicitatieverplichting, dat wil zeggen relatief kostbaar.

Draagvlak in het veld
Als bij de sollicitatieverplichting, dat wil zeggen goed.

Privacybescherming
Als bij de sollicitatieverplichting, dat wil zeggen zonder onoverkomelijke problemen.

Aselecte controles

Validiteit
Deze verplichting wordt gekenmerkt door een specifiek medisch karakter. Bovendien is er sprake van een zeer subjectieve beoordeling van ‘zich beter voelen’, waarna dit gemeld moet worden aan de uitkeringsinstantie. Aselecte controles spitsen zich hier in de praktijk niet op toe.
In theorie is het echter mogelijk om artsen en arbeidsdeskundigen te laten controleren in hoeverre een arbeidsgehandicapte cliënt hersteld is, zonder dit te melden. Voor een (belangrijk) deel zal echter door de cliënt zelf tijdens de controle aangegeven moeten worden of hij zich beter voelt dan voorheen. Bovendien betreft een controle een momentopname, waardoor het lastig is om vast te stellen wanneer de verbetering is ingetreden, of de eventueel verbeterde medische situatie blijvend is of slechts van korte duur.
Deze complicaties maken het risico dat bij controles regelovertredingen over het hoofd kunnen worden gezien, aanzienlijk.

Betrouwbaarheid
Omdat de controlerende instantie zelf kan bepalen hoeveel controles uitgevoerd worden, is het relatief eenvoudig om de betrouwbaarheid van de uitkomsten desgewenst te vergroten. Het validiteitsrisico is daarmee echter niet verkleind.

Doorlooptijd
Geen bijzonderheden ten opzichte van aselecte controles bij andere nalevingsverplichtingen.

Kosten
Uitgebreide medische controles zijn arbeidsintensief en brengen relatief hoge kosten met zich mee.

Draagvlak in het veld
In het veld worden geen aselecte controles gebruikt om vast te stellen of een cliënt eventueel herstel niet heeft gemeld. Gezien de omvangrijke problemen ten aanzien van objectiveerbaarheid, ziet men hier het nut niet van in.

Privacy
Geen bijzonderheden ten opzichte van aselecte controles bij andere nalevingsverplichtingen.

Administraties

Klanten waarvan het UWV vaststelt dat zij de verplichtingen niet voldoende naleven, komen via CDF tot uiting. Voor de klanten die niet zijn onderzocht, kunnen echter geen directe uitspraken worden gedaan. Generalisatie moet hier uitkomst bieden.
Daarbij moeten we rekening houden met de *controlekans, detectiekans en sanctiekans*. Alles wat hierover is gezegd in paragraaf 4.2.4 is ook hier van toepassing.

Validiteit en betrouwbaarheid
Op grond van het bovenstaande beoordelen we de validiteit en betrouwbaarheid als beperkt.

Doorlooptijd
Geen specifieke opmerkingen.

Kosten
Geen specifieke opmerkingen.

Draagvlak in het veld
Geen specifieke opmerkingen.

Privacybescherming
Geen specifieke opmerkingen.

5.6.3 Conclusie

Geen van de beschreven methodes levert resultaten met een behoorlijke validiteit, maar de zelfrapportage scoort verhoudingsgewijs nog het best.

6 AFSLUITING

6.1 Inleiding

In dit slothoofdstuk zetten we de methodieken op een rij, vergelijken hun kwaliteiten en komen tot een beoordeling. We destilleren hieruit een advies voor de toekomstige inrichting van POROSZ.

6.2 Overzicht van prestaties van methoden naar verplichting

In de voorgaande hoofdstukken zijn de prestaties van de onderscheiden methoden beschreven. De bevindingen laten zich samenvatten in onderstaande tabel 6.1. Bij de beoordeling hanteren we vier waarden: slecht, matig, redelijk en goed. Alleen voor de kosten hanteren we andere waarden: gunstig en ongunstig.

Tabel 6.1 Beoordelingsmatrix

		Zelfrapportage		Aselecte controle	Administratie
		RR	directe bevraging		
1	**validiteit**	**B**	**C**	**C**	**D**
	sollicitatie	B	B	C	D
	traject	B	B	C	D
	inkomen zwart	B	C	C	D
	inkomen wit	B	C	C	A
	gezamenlijke huishouding	B	C	B	D
	herstel	D	D	D	D
2	**betrouwbaarheid**	**C**	**B**	**C**	**B**
	sollicitatie	C	B	C	A
	traject	C	B	C	A
	inkomen zwart	C	B	C	-
	inkomen wit	C	B	C	A
	gezamenlijke huishouding	C	B	C	B
	herstel	C	C	C	B
3	**bereik**	**A**	**B**	**C**	**D**
4	**doorlooptijd**	**A**	**A**	**A**	**D**
5	**kosten**	**ongunstig**	**ongunstig**	**ongunstig**	**gunstig**
6	**draagvlak**	**B**	**B**	**C**	**D**
7	**privacy**	**C**	**C**	**C**	**C**

A = goed B = redelijk C = matig D = slecht

Aangezien er op de punten van validiteit en betrouwbaarheid de nodige variatie bestaat tussen de verschillende verplichtingen, zijn deze nader gespecificeerd.

6.3 Vergelijking van de methodes: bespreking

Validiteit. Alle methodes (administratie, controle en zelfrapportage) kampen met validiteitsproblemen:

- Administraties; omdat veel vormen van niet-naleving er niet in terechtkomen.
- Controles; omdat voor veel vormen van niet-naleving in de praktische uitvoering eenvoudigweg niet gecontroleerd kan worden en omdat sommige begrippen moeilijk te operationaliseren zijn.
- Zelfrapportage; omdat overtreders niet allemaal mee willen werken, de vragen niet begrijpen of geen eerlijke antwoorden geven.

Alles afwegende zijn naar ons oordeel de validiteitsproblemen bij zelfrapportage het best oplosbaar. Op het punt van validiteit onderscheidt de RR-methode zich nog ten opzichte van de directe bevraging, als het gaat om onderwerpen die gevoelig liggen bij de respondenten.
Tot de voorwaarden behoort dat gebruik wordt gemaakt van een adequaat steekproefkader, zoals bijvoorbeeld de administratie van de Belastingdienst, het UWV, de SVB.

Betrouwbaarheid: Als het gaat om precisie van onderzoeksresultaten is de administratieve benadering beter dan beide andere benaderingen. Dat wordt voornamelijk veroorzaakt door het ontbreken van de betrouwbaarheids-marges, die horen bij het gebruik van steekproeven in de beide andere methodieken.

Bereik. De zelfrapportage, en dan met name de RR-methode, is beter geschikt voor bevraging van het hele amalgaam aan verplichtingen dan de andere methoden. Immers, deze methode levert behoorlijk valide resultaten op alle nalevingsverplichtingen.
De administratieve benadering en de aselecte controle scoren beiden op onderdelen slecht. Zo kan met de administratieve benadering de naleving van de sollicitatieverplichting en de trajectverplichting niet goed worden waargenomen. Aselecte controle lijkt redelijk te werken voor het waarnemen van verzwijgen van gezamenlijke huishouding, maar veel minder voor de overige verplichtingen. Ook van belang is dat aselecte controle een heel ander soort waarneming doet, namelijk van het ‘hier en nu’. Dit verschilt van de huidige bevraging in POROSZ en heeft ook gevolgen in die zin dat het een veel lagere prevalentie zal opleveren.

Doorlooptijd. Op dit punt hebben alle drie de methodieken hun eigen karakteristieken, die uiteindelijk niet zoveel verschillen lijken op te leveren in doorlooptijd, maar wel in de mogelijkheid om in februari van het jaar t+1 te kunnen rapporteren.
Voor de zelfrapportage geldt als belangrijk kenmerk dat het steekproefkader beschikbaar moet zijn. Als gebruik wordt gemaakt van panels, zoals bij de laatste uitvoering van POROSZ, dan is dit verder geen knelpunt. Als echter gebruik wordt gemaakt van bijvoorbeeld de Belastingdienst, dan kan dit een probleem worden, omdat de Belastingdienst omstreeks mei-juni pas de gegevens over voorgaand belastingjaar beschikbaar kan stellen.
Voor de aselecte controle geldt dat de controles vaak arbeidsintensief zijn en om die reden nogal wat tijd kunnen vergen. Met deze beide methodes kunnen bevindingen over verslagjaar 2006 worden geleverd in februari 2007.
Dat geldt niet voor de administratieve benadering, immers de bestanden over 2006 moeten beschikbaar zijn voordat met de analyse kan worden begonnen. Het is zeer onwaarschijnlijk dat dit lukt voor februari 2007.

Kosten. Kosten zijn natuurlijk sterk afhankelijk van de definitieve specificaties voor en inrichting van het onderzoek. Wel zijn er globale indicaties te geven. Zelfrapportage met behulp van randomized response en aselecte controle behoren tot de duurdere methoden. De administratieve benadering is ten opzichte van deze methoden voordeliger.
Het is overigens niet gezegd dat een directe bevraging (veel) goedkoper uit zal pakken dan een randomized-responseaanpak. Vanwege de beperktere waarneming van regelovertreding ligt de prevalentie over het algemeen lager en zijn de betrouwbaarheidmarges kleiner. Dientengevolge zullen ook de veranderingen daarin kleiner zijn. Om die kleine veranderingen met directe bevraging te kunnen waarnemen, zijn weer grotere steekproefaantallen nodig. Welk van de twee varianten duurder uitvalt, is op voorhand niet aan te geven. Dat is namelijk afhankelijk van de verschillen in prevalenties tussen een directe bevraging en een randomized-responseaanpak.

Draagvlak in het veld. Bij de SVB zijn goede ervaringen opgebouwd met het vaststellen van nalevingsniveaus door middel van aselecte controle. Men is van plan dat ook de komende jaren voort te zetten en ziet daarbij geen toegevoegde waarde in onderzoek op basis van zelfrapportage. Als de kwaliteit van POROSZ verbetert, kan een dergelijke meerwaarde ontstaan en is het denkbaar dat deze methodiek de plaats overneemt van de aselecte controles.
De administratieve benadering spreekt partijen aan vanwege het arbeidsextensieve karakter en de kosten. Tegelijkertijd aarzelt men over de validiteit. Onder de partijen is er draagvlak voor de zelfrapportage, ook op basis van randomized response. Wel leven er ideeën voor verbetering.

Privacy
Bij de uitvoering van POROSZ moet terdege rekening worden gehouden met de bepalingen in de Wet Bescherming Persoonsgegevens (WBP), alsmede

met verwante bepalingen. Onoverkomelijke hindernissen zijn ons niet gebleken, voor geen van de methoden.

6.4 Conclusie

Wij stellen vast dat voor zowel de aselecte controle (knelpunten: validiteit, bereik) als de administratieve benadering (knelpunten: validiteit, bereik en doorlooptijd) geldt dat er obstakels zijn om hen te gebruiken als primaire methodiek voor POROSZ. Zelfrapportage is ook niet zonder problemen, maar levert betere prestaties als het gaat om cruciale aspecten als validiteit én bereik. Bij gevoelige vraagitems is validiteitswinst te verwachten van de RR-methode ten opzichte van de directe bevraging.

6.5 Uitgangspunten bij opstellen advies

Bij het formuleren van ons advies hanteren we de volgende uitgangspunten:

i. We streven naar het hanteren van één methode om alle nalevingsverplichtingen mee te onderzoeken. Dat leidt tot een overzichtelijk onderzoeksdesign, hetgeen bevorderlijk is voor de transparantie en de kosten- en risicobeheersing. Het bevordert de begrijpelijkheid voor de lezer en bovendien bevordert het de vergelijkbaarheid van de resultaten (tussen verplichtingen, tussen regelingen). We benadrukken dat het hier gaat om een streven, niet meer en niet minder. Het is dus geen keurslijf, maar geeft richting. De mate waarin een methode presteert op dit onderwerp komt tot uiting in het beoordelingscriterium *bereik*.
ii. Naast het *bereik* wegen we de prestaties op het criterium *validiteit* zwaarder dan de overige criteria, omdat juist hierdoor de kwaliteit van het onderzoeksresultaten wordt bepaald. Betrouwbaarheid speelt daarin zeker ook een rol, maar wij achten die ondergeschikt aan de validiteit.[1] Het maximeren van dit laatste is het eigenlijke doel van ons advies. Goed denkbaar is dat een beleidsoptiek leidt tot een andere weging en daarmee een andere uitkomst. Alle ingrediënten om dit te doen, zijn in dit verslag opgenomen.

6.6 Advies POROSZ 2006 en verder

Dit leidt ons tot het volgende advies voor de opzet van POROSZ, dat betrekking heeft op de methode, het steekproefkader en de operationalisering van begrippen.

[1] Ter toelichting: een valide onderzoeksresultaat met beperkte betrouwbaarheid (ruime marges) is veel beter generaliseerbaar dan een resultaat met een goede betrouwbaarheid (geringe marges) en een beperkte validiteit.

Methode

a. Zelfrapportage als primaire methode voor gegevensverzameling
We stellen voor om POROSZ primair te (blijven) stoelen op de methode van zelfrapportage. Met deze methode kunnen alle verplichtingen worden onderzocht, dus het bereik is goed en de validiteit is acceptabel, dat wil zeggen, ten minste beter dan bij de alternatieve methoden.

Daarbij dient zich de keuze aan tussen de RR-variant en de directe bevraging. De directe bevraging heeft als belangrijkste voordelen een eenvoudigere opzet van het vraaggesprek, een betere begrijpelijkheid voor de respondent en de mogelijkheid tot gebruik van selectievragen.
Het is overigens niet gezegd dat een directe bevraging veel goedkoper uit zal pakken dan een randomized-responseaanpak. Vanwege de beperktere waarneming van regelovertreding ligt de prevalentie over het algemeen lager en zijn de betrouwbaarheidmarges kleiner. Dientengevolge zullen ook de veranderingen daarin kleiner zijn. Om die kleine veranderingen met directe bevraging te kunnen waarnemen, zijn weer grotere steekproefaantallen nodig. Welk van de twee varianten duurder uitvalt, is op voorhand niet aan te geven. Dat is namelijk afhankelijk van de verschillen in prevalenties tussen een directe bevraging en een randomized-responseaanpak. Alleen de empirie kan hier uitkomst bieden. Dat houdt in dat in een volgende meting beide methoden toegepast zouden moeten worden om die verschillen vast te stellen.
Een aanvullend probleem is dat bij gebruik van directe bevraging de kans groter is dat, in geval van een verplichting die niet veel wordt overtreden, de waargenomen prevalentie op nul uitkomt.

Om de keuze tussen RR en directe bevraging te kunnen onderbouwen, is het nodig om te onderzoeken welke validiteitswinst wordt geboekt met het hanteren van de RR-variant en tegen welke kosten. Ook moet SZW bepalen hoe erg het is als de prevalentie van een de verplichtingen op nul uitkomt. Dat pleit voor een (voorlopig) eenmalige meting waarin beide methodieken worden toegepast. Die meting geeft antwoord op de volgende vragen (naast de reguliere onderzoeksvragen):

- Wat is per nalevingsniveau de validiteitswinst van de RR-aanpak ten opzichte van directe bevraging?

- Wat zijn de benodigde steekproefaantallen per nalevingsniveau bij een RR-aanpak en een directe bevraging om veranderingen in de tijd te kunnen vaststellen?

Afhankelijk van de antwoorden op deze vragen kan POROSZ daarna definitief worden vormgegeven.
Blijkt de validiteitswinst van de RR-aanpak beperkt te zijn en zijn de benodigde steekproefaantallen voor een directe bevraging veel kleiner dan die voor een RR-aanpak, dan ligt het voor de hand voor de vormgeving van POROSZ voor een design te kiezen waarbij directe bevraging het uitgangspunt vormt en bij een beperkt aantal metingen een RR-aanpak wordt gebruikt om de verschillen

tussen beide methodieken te onderzoeken en de directe bevraging daarmee te ijken.
Blijkt de validiteitswinst van de RR-aanpak groot en de verschillen in benodigde steekproefaantallen gering, dan ligt het voor de hand om de RR-aanpak te blijven continueren.

b. Beide andere methoden secundair, gericht, toepassen
Bestaande aselecte controleactiviteiten van de SVB, het UWV en enkele gemeenten gebruiken om:

- de non-respons onder a. nader te kunnen duiden;
- te bezien of deze, in geval van de AOW- en Anw-prestatie-indicator 'naleving opgave gezamenlijke huishouding', in de plaats kunnen treden van POROSZ. Daarvoor is nadere analyse van overeenkomsten en verschillen tussen POROSZ en het genoemde SVB-onderzoek nodig.[2]

Denkbaar is dat ook bestaande statistieken een ondersteunende rol kunnen spelen. Het idee daarbij is dat niet het absolute niveau van niet-naleving wordt gebruikt om een eventueel non-responsprobleem in kaart te brengen, maar de verhoudingen van niet-naleving tussen te onderscheiden groepen (leeftijd, geslacht, stad/platteland et cetera.).

Ook kunnen zij worden benut om een ondergrens van de prevalentie in kaart te brengen. In het geval van verzwijgen van witte inkomsten kan dat met administraties. In geval van samenleeffraude op basis van aselecte controles, op voorwaarde dat de bevindingen op een behoorlijke manier te generaliseren zijn naar de populatie.

Steekproefkader

c. Gebruik van een ander steekproefkader
Als resultaten uit de steekproef dicht moeten liggen bij de werkelijke waarden, is de kwaliteit van de steekproef van groot belang. In verband met het kunnen generaliseren van de onderzoeksresultaten is het nodig dat het steekproefkader vastomlijnd en bekend is, zodat een oordeel over de kwaliteit van de respons gegeven kan worden. Het huidige steekproefkader (access pools) voldoet wat ons betreft niet aan die voorwaarde, met name vanwege de onduidelijke representativiteit.
Wij stellen voor om steekproeven te trekken uit de administraties van de Belastingdienst (bijstandsontvangers, eventueel ook WW en WIA), van het UWV (WW en WIA) en, indien aan de orde, de SVB (AOW, Anw, AKW).
Een bijkomend voordeel van gebruik van de administraties van het UWV is dat hieruit ook kan worden afgeleid of er een trajectaanbod geldt.

[2] Dat kan door terug te blikken op de resultaten over 2004, of dit te doen met de nog te produceren resultaten over 2006.

d. Niet-Nederlandssprekenden op volwaardige wijze betrekken in het onderzoek[3]

Als gevolg van praktische overwegingen zijn tot op heden mensen die niet of slecht Nederlands spreken, niet vertegenwoordigd in de steekproeven. Daarvoor bestaat geen inhoudelijke grond. Dit verbetert 'vanzelf' met de gebruikmaking van het voorgestelde steekproefkader. Dit stelt wel nadere eisen aan vertalingen van vragenlijsten en/of overwegingen ter zake van *multi channeling.*

Operationalisering

e. Nadere uitwerking van de definitie van regelovertreding

Ter verbetering van de interpretatiemogelijkheden van de resultaten is het verstandig om de periode waarop regelovertreding betrekking kan hebben te standaardiseren (bijvoorbeeld: één jaar, een halfjaar, drie maanden).

f. Wijziging van de constructie van de prestatie-indicatoren

Tot op heden is de mate van regelovertreding gerelateerd aan de totale groep van uitkeringsontvangers, dus inclusief personen die de desbetreffende regel niet kunnen overtreden. Dat komt wel tegemoet aan de praktische uitvoering van het onderzoek, maar is wat ons betreft inhoudelijk niet erg zuiver en we geven in overweging om dat bij te stellen. De eerlijkheid gebiedt te zeggen dat bijstelling voor de uitvoering van het onderzoek grote gevolgen kan gaan hebben. Óf de opdrachtgever zou moeten accepteren dat voor specifieke nalevingsniveaus met fors grotere steekproeven gaat worden gewerkt, óf de opdrachtgever moet bereid zijn het meten van bepaalde nalevingsniveaus niet meer op te nemen. Op dit moment is niet bekend (vanwege onbekendheid met de omvang van de te maken selecties) wat de noodzakelijk aanpassingen zouden zijn. Alleen met behulp van een pilot (die bijvoorbeeld meeloopt in het komende POROSZ-onderzoek) waarmee de benodigde empirische gegevens kunnen worden verzameld, is een onderbouwd antwoord mogelijk.

g. Nadere selectie prestatie-indicatoren.

Als gevolg van f daalt het aantal personen waarover een uitspraak kan worden gedaan, hetgeen leidt tot grotere betrouwbaarheidsmarges, althans bij gelijkblijvende steekproefomvang en -inrichting.

Gevolg is dat óf genoegen moet worden genomen met deze verminderde precisie, of de steekproefinrichting en/of -omvang moet worden bijgesteld. Zo kan de tamelijk kleine groep WW'ers met trajectverplichting gericht worden benaderd door een steekproef te trekken uit de trajectadministratie van UWV. Ook verdient het overweging om de selectie van prestatie-indicatoren te beperken tot uitsluitend die verplichtingen die voor een groot deel van de populatie gelden.

[3] Wij benadrukken nogmaals dat het onderzoek en het advies geen betrekking hebben op in het buitenland wonende uitkeringsontvangers.

LITERATUUR

Bockenholt, U. en P.G.M. van der Heijden (2004). *Randomized-Response Models for Measuring Noncompliance: Risk-returns, Social Influences and Self-protectiveness.* Ongepubliceerd manuscript ter beschikking gesteld door auteurs.

Boeije, H.R. en G. Lensvelt-Mulders (2002). Honest by chance: A qualitative interview study to clarify respondents' (non-)compliance with computer assited randomized response in *Bulletin Methodologie Sociologique*, 75: 24-39.

Feinstein, J.S. Detection Controlled Estimation. In: *Journal of Law and Economics*, Vol. 33, No. 1 (Apr., 1990) , pp. 233-276.

Jablonowski, C.J. Organizational, operational and natural hazards in the upstream oil and gas industry. Pensylvania, 2002.

Lensvelt-Mulders, G.J.L.M., A. van den Hout en P.G.M. van der Heijden (2004). An Experimental validation of a Computer-Assited Randomized response Survey' paper Universiteit Utrecht.

Lensvelt-Mulders, G.J.L.M., J.J. Hox, en H.R. Boeije (2003). *Evaluating Compliance with the Computer Assisted Randomized Response technique. A Qualitative research in the background of lying and cheating*, paper Universiteit Utrecht, 8 januari 2003.

Lensvelt-Mulders, G.J.L.M., J.J. Hox, P.G.M. van der Heijden en C.J.M. Maas (2004). Meta-Analysis of Randomized-Response Research, *Sociological Methods and Research*, vol 33, nov. 2004, p. 1-30.

Lensvelt-Mulders, G.J.L.M., P.G.M. van der Heijden en O. Laudy (2006 nog te verschijnen). A validation of a computer-assisted randomized response survey to estimate the prevalence of fraud in social security, *Journal of the Royal Statistal Society*, vol. 169.

Stoop, I.A.L. (2005). *The hunt for the last respondent. Non-response in sample surveys*. Den Haag: SCP.

Soethout, J.E. en R.C. van Waveren (2003). *Eindmeting evaluatie LAT-project*. Amsterdam: Regioplan Beleidsonderzoek.

Voogt, R.J.J. (2004). *I'm not interested. Non response bias, Response bias and Stimulus Effects in Election Research*. Proefschrift Universiteit van Amsterdam.

BIJLAGEN

BIJLAGE 1

INFORMANTEN

Ten behoeve van dit onderzoek is gesproken met:

- CBS, Henk Amptmeijer, Wim Schaasberg, Harold Eding
- Belastingdienst, Harry van Berkel
- Gejo Duinkerken
- SVB, Frits Werner, Paul Burroughs
- UWV, Harry Aukes, Margreet Stoutjesdijk, Susanne Rietbroek
- Inlichtingenbureau, Peter Jansz
- SIOD, Nico Babbeko
- Gemeente Amsterdam, Dienst Persoonsgegevens, Jan Brouwer
- IOPS, Peter van der Heijden, Ger van Gils
- NISSO
- WODC
- TNS NIPO
- NIVEL

En met vertegenwoordigers van enkele gemeenten, afdeling SoZaWe.

BIJLAGE 2

Tabel B2-1 Nalevingsniveaus naar onderscheiden verplichtingen

		2003	2004	2005	2006	2007	2008	2009
Re-integratie								
23.01.03	nalevingsniveau van deelname aan een traject	-	-	-		-		
WBB								
30.01.01	nalevingsniveau opgave van inkomsten	0.61	0.64	0.67				0.73
30.02.03	nalevingsniveau sollicitatieverplichtingen	0.68	0.7	0.72				0.8
30.02.04	nalevingsniveau opgave van inkomsten uit arbeid	0.74	0.75	0.76				0.8
WW								
	nalevingsniveau opgave van inkomsten		0.75	0.75	>0.75	>0.75		>0.75
	nalevingsniveau verplichting werk te vinden		0.66	0.66	>0.66	>0.66		
	nalevingsniveau opgave van inkomsten uit arbeid							
WAO/WAJONG/WAZ								
31.01.01	nalevingsniveau verplichtingen i.h.k.v. herstel	0.76	0.77	0.78				0.82
31.01.02	nalevingsniveau opgave inkomsten uit arbeid	0.78	0.79	0.8				0.84
Anw								
32.01.01	nalevingsniveau melden van gezamenlijke huishouding	onbekend						
32.01.02	nalevingsniveau opgave inkomen en wijziging daarvan	onbekend						
AOW								
32.02.01	nalevingsniveau melden van gewijzigde samenlevingsvorm	onbekend						
32.02.02	nalevingsniveau opgave inkomen en wijziging daarvan van partner jonger dan 65	onbekend						
AOW								
33.02.01	nalevingsniveau tijdige opgave wijziging in- en uitwonend	onbekend						
33.02.02	nalevingsniveau opgave inkomsten kinderen 16 jaar en ouder	onbekend						