

Samenvatting

Inleiding

In het Hoofdlijnenakkoord van het tweede kabinet Balkenende is vastgelegd dat het illegaal verblijf door vreemdelingen in Nederland krachtiger dan eerder zal worden tegengegaan. In Nederland illegaal verblijvende vreemdelingen zijn een voorbeeld van een verborgen populatie. Om te kunnen nagaan of het voorgenomen beleid effect heeft is het dan ook nodig om de omvang van dit soort verborgen populaties te kunnen schatten. Deze studie geeft een overzicht van methoden die in het algemeen gebruikt kunnen worden om uitspraken te doen over de omvang van verborgen populaties. Daarbij wordt in het bijzonder ingegaan op de geschiktheid van de methoden voor het bepalen van de omvang van de populaties van illegale vreemdelingen en het schatten van de omvang van deelpopulaties hiervan, bijvoorbeeld naar nationaliteit of woonplaats.

Formeel kunnen populaties verborgen zijn om twee redenen:

- Er bestaat geen lijst waarop alle leden van de populatie zijn vermeld, geen steekproefkader dus. Waar het gaat om illegalen zijn er natuurlijk wel lijsten waarop een aantal van deze personen voorkomen, bijvoorbeeld een medische registratie, politiebestanden, leerlingenbestanden van scholen en dergelijke, maar er is geen complete lijst waarop elke illegaal is vermeld.
- Er is wel een lijst waarop alle leden van de populatie zijn vermeld, maar er staan veel meer personen op die lijst en op voorhand kan niet worden gezegd wie tot de verborgen populatie behoort. In dit geval spreken we pas van een ‘verborgen populatie’ wanneer dit vaststellen moeilijk is. De vraag of iemand van baan wil veranderen is gemakkelijk gesteld en beantwoord, maar het is veel moeilijker om valide antwoorden te krijgen op medische vragen of vragen naar sociaal onwenselijk of strafbaar gedrag.

De omvang van de verborgen populatie is vaak een gegeven dat slechts met veel moeite kan worden achterhaald. Daarom is het van belang om een goed inzicht te hebben in alle beschikbare alternatieven om dit doel te bereiken. De inhoud van dit rapport is vooral gebaseerd op literatuuronderzoek; op sommige punten klinkt echter ook de praktische ervaring door die de auteurs met deze materie hebben.

Bij de definitie van een illegaal zijn twee invalshoeken mogelijk: de economische en de demografische. Volgens de economische invalshoek is een illegaal een vreemdeling die arbeid verricht zonder geldige werkvergunning. Bij de demografische definitie staat echter de verblijfsvergunning centraal. Een illegaal is dan iemand die zonder geldige verblijfsvergunning in Nederland verblijft. Het doet er niet toe of deze persoon legaal of illegaal het land is binnengekomen. Dit is de definitie die in het overgrote deel dit rapport wordt gehanteerd.

Het rapport is geschreven vanuit een aantal verschillende doelstellingen. In de eerste plaats is het een overzicht van statistische methoden om in het algemeen de omvang van verborgen populaties te schatten. In die zin is dit rapport bedoeld als een naslagwerk. Dit overzicht wordt gegeven in hoofdstuk 2, in een aantal afzonderlijke paragrafen. Daarbij is getracht om niet alleen de technische werking van de methoden uit te leggen maar een gevoel te geven van hoe zich dit verhoudt met de *state of the art* van de statistische methodologie. Wie na het lezen van zo'n paragraaf de methode in een lezing in een buitenlands congres noemt zal niet worden verrast door voor hem totaal nieuwe inzichten die inmiddels tot de geaccepteerde wetenschap behoren. In de tweede plaats beoogt het rapport inzicht te geven in welke methoden zich lenen voor het tellen van illegalen. Niet elke methode die gebruikt wordt voor het schatten van verborgen populaties is daarvoor geschikt. Maar voor elke methode geldt wel dat daarbij wordt uitgegaan van een zekere geïdealiseerde werkelijkheid. De echte werkelijkheid voldoet daar vaak niet aan. Daarom is in hoofdstuk 3 een overzicht gegeven van een aantal praktijkvoorbeelden in binnen- en buitenland. Dan blijken de methoden wel bruikbaar te zijn, maar altijd weer te moeten worden aangevuld met specifieke aanpassingen. De derde doelstelling is daarmee om de lezer een gevoel te geven van wat er bij komt kijken om een illegalentelling in de praktijk uit te voeren. Ten slotte, en hiervoor is hoofdstuk 4 bedoeld, biedt het rapport ondersteuning bij het kiezen voor een specifieke methode die bij een illegalenmonitor kan worden gebruikt.

De methoden

De methoden die in dit rapport worden besproken lopen sterk uiteen naar aard, vooronderstellingen, gegevensbehoefte. Met sommige is reeds ervaring opgedaan bij het schatten van illegalen, andere zijn gebruikt voor andere verborgen populaties, maar niet voor illegalen, en ook zijn er methoden waarvan het gebruik wel denkbaar is, maar waarvoor praktische ervaring nog ontbreekt.

Vangst-hervangstmethoden

Een verzameling van methoden waarmee juist in Nederland al ruime ervaring is opgedaan zijn de *vangst-hervangstmethoden*. Dit zijn manieren om steekproefonderzoek te doen zonder steekproefkader. Steekproefelementen worden letterlijk 'gevangen', maar ook weer losgelaten. Door verschillende steekproeven te trekken en te bepalen hoeveel elementen in meer dan één steekproef voorkomen kan een schatting worden gemaakt van de omvang van de populatie waaruit de steekproeven zijn getrokken. De naam 'vangst-hervangst' (capture-recapture)

komt oorspronkelijk uit de biologie waar deze methode wordt gebruikt bij het schatten van populatiegroottes van diersoorten. Hierbij wordt gepoogd de dieren herhaaldelijk te vangen. In de biologie is het gebruikelijk om de steekproeven op verschillende momenten in de tijd te trekken. Letterlijk worden dieren gevangen en later weer hervangen. Maar het is ook denkbaar dat twee vangsters onafhankelijk van elkaar dieren vangen en op die manier tot steekproeven komen die min of meer gelijktijdig tot stand zijn gekomen. Deze metafoor is gewoonlijk van toepassing in de epidemiologie en bij overheidsstatistieken. Men maakt dan gebruik van bestaande registraties waarin mensen zijn 'gevangen'. Na koppeling kunnen de aantallen personen worden bepaald die, al dan niet gelijktijdig, in de registraties voorkomen. Onder bepaalde veronderstellingen kan dan het aantal personen worden geschat dat in geen van de gekoppelde registraties voorkomt. Tezamen leveren de personen uit het gekoppelde bestand en de niet-voorkomende personen de schatting van de omvang van de populatie op.

Gebruik van gekoppelde registraties

Om vangst-hervangstmethoden op basis van gekoppelde registraties toe te kunnen passen op de illegale populatie moet aan een veronderstelling zijn voldaan, te weten:

- een gesloten populatie: gedurende de periode waarop de schatting betrekking heeft moet de populatie constant blijven
- geen koppelfouten: de personen worden correct en uniek geïdentificeerd
- opname in bestanden onafhankelijk: de kans om in alle bestanden voor te komen is het product van de kansen om in elk van de bestanden voor te komen.

Aan deze voorwaarden is niet altijd voldaan. Over het algemeen is wel bekend hoe het niet voldaan zijn aan de veronderstellingen de schattingen beïnvloedt, met name of dit tot over- of onderschattingen leidt. Ten slotte is er een vierde voorwaarde, die meer het karakter van een wens dan van een eis heeft:

- de kans om in een registratie voor te komen is identiek voor alle leden van de verborgen populatie

Wanneer hier niet aan voldaan is, zijn er statistische methoden beschikbaar om met deze verschillen in kansen om te gaan. Ze moeten dan wel in het model worden meegeschat, hetgeen de statistische berekening van de omvang van de verborgen populatie compliceert, maar niet onmogelijk maakt.

Herhaald vangen

Wanneer er sprake is van herhaald vangen (aanhouden van illegalen), dan geschiedt het schatten van de omvang van de populatie op basis van één enkel bestand, waar de illegalen meerdere keren in kunnen voorkomen. Hiervoor wordt een theoretische verdeling, de Poissonverdeling, gebruikt, die aangeeft wat de kans is dat iemand 0, 1, 2 enzovoorts aantal keren wordt gevangen. Deze verdeling kent een parameter die op grond van de gegevens wordt geschat, waarna de kans op het niet voorkomen in het bestand kan worden bepaald, en daarmee dus de totale omvang van de populatie. De meest cruciale veronderstelling hier is dat deze verdeling daadwerkelijk gebruikt mag worden. Impliciet wordt daarbij verondersteld dat de kans op een vangst niet verandert doordat men niet gevangen wordt, en de kans op een hervangst niet beïnvloed wordt door het feit dat men gevangen is. Een veronderstelling, waar men net als bij de variant gebaseerd op gekoppelde bestanden met zwaar statistisch geschut wel omheen kan is dat iedere illegaal dezelfde Poissonparameter heeft. Wanneer hieraan niet is voldaan kunnen de verschillen in parameters in het model worden meegenomen.

Sneeuwbalsteekproeven

Ook sneeuwbalsteekproeven zijn steekproeven zonder steekproefkader. Bij sneeuwbalsteekproeven wordt gestart met een kleine initiële steekproef, waarna vervolgens het trekken van nieuwe elementen wordt overgelaten aan de respondenten die al in de steekproef zaten. Er zijn twee varianten van dit soort steekproeven: (i) de 'klassieke' sneeuwbalsteekproef, waarbij de respondenten aan de onderzoekers de namen opgeven van potentiële nieuwe respondenten en (ii) *respondent driven sampling*, waarbij de respondenten zelf potentiële nieuwe respondenten benaderen met het verzoek om aan het onderzoek deel te nemen. Deze laatste methode heeft het voordeel dat niemand de identiteit van een ander prijsgeeft. In het kader van onderzoek naar illegalen is dit een aanzienlijk voordeel, want uit eerder onderzoek is gebleken dat er een grote angst is om deel te nemen. Het is zelfs de vraag of deze angst via respondent driven sampling weggenomen kan worden. Deze vraag kan niet worden beantwoord, omdat de methode nooit in deze context is toegepast. Het maken van schattingen op basis van uitkomsten van sneeuwbalsteekproeven is statistisch complex. Voor de klassieke methode zijn er alleen handzame formules beschikbaar ingeval er sprake is van een startsteekproef en één golf van nieuwe respondenten. In het geval dat het aantal illegalen als een percentage van een groter geheel worden berekend (bijvoorbeeld alle allochtonen) is er wel een meer algemene methode beschikbaar, waarbij echter moet worden verdisconteerd dat mensen met

een groot sociaal netwerk zijn oververtegenwoordigd in de steekproef. Dit maakt het schattingsproces complex.

Detection Controlled Estimation

Detection Controlled Estimation (DCE) is een schattingstechniek om simultaan kenmerken van overtreeders en hun opspoorders te schatten. Daarbij kan het gaan om personen, bijvoorbeeld studenten die mogelijk frauderen bij een examen, of bedrijven die de milieuregels overtreden. In beide gevallen is er sprake van een controleur. Bij de examens is het de surveillant, bij bedrijven de opsporingsambtenaar. In beide gevallen is er sprake van cases van paren: aan elke potentiële overtreder is een controleur gekoppeld. Een overtreding wordt alleen maar geregistreerd wanneer deze zowel gemaakt wordt door de potentiële overtreder als ontdekt wordt door de controleur. Niet ontdekte overtredingen worden dus niet geregistreerd. Het schatten van de omvang van de verborgen populatie komt daarmee neer op het bij-schatten van de niet ontdekte overtredingen. De kans op het maken van een overtreding noemen we p_1 ; de kans op ontdekking van een gemaakte overtreding gelijk is aan p_2 . De kans op een geregistreerde overtreding is dan gelijk aan $p_1 p_2$. De kans dat dit niet het geval is, is $1 - p_1 p_2$. In dit laatste geval is dus niet bekend of er een overtreding heeft plaatsgehad, alleen dat deze niet is geconstateerd. De kansen op overtreding en op opsporing worden met behulp van de techniek gekoppeld aan de kenmerken van de overtreeders en van de controleurs. Wanneer deze relatie eenmaal geschat is wordt het ook mogelijk om uitspraken te doen over het aantal niet opgespoorde overtredingen. In dit kader kan de overtreding worden geïnterpreteerd als 'illegaal zijn'. De controleur kan inspecteur van een opsporingsinstantie zijn, maar ook een respondent in een onderzoek, die rapporteert of hij weet of mensen in de buurt illegaal in Nederland verblijven. DCE is een zeer flexibele econometrische techniek, die allerlei verfijningen in de modelformulering toelaat. Zo kan er sprake zijn van voorkennis bij controleurs op strategisch niveau (men weet de kenmerken van overtreeders), op tactisch niveau (men ziet overtredingen aankomen) en er kan de mogelijkheid worden opgenomen dat overtreeders zichzelf aangeven. De prijs die voor dit soort ad hoc aanpassingen wordt betaald is de noodzaak van geavanceerd programmeerwerk, en dus een lange doorlooptijd, om de schattingsmethoden uit te voeren.

Schattingen op basis van postcodegebied

Nederland is ingedeeld in een fijnmazig netwerk van postcodegebieden. Van deze postcodes zijn op het meest gedetailleerde niveau (6 digits) kenmerken bekend. Uiteraard zijn van deze kenmerken ook de landelijke verdelingen bekend. Dit maakt het mogelijk om het postcodebestand

als steekproefkader te gebruiken. Bij verborgen populaties die sterk gelieerd zijn aan bepaalde postcodegebieden, of aan postcodegebieden van een bepaald type, biedt dit de mogelijkheid om via postcodevariabelen tot landelijke omvangschattingen te komen. Voor de hand liggende populaties zijn daklozen, drugsverslaafden, prostituees en met name ook illegalen, omdat in de vreemdelingenadministratie wordt vastgelegd in welk postcodegebied hun woonadres is. Zo zijn er in de periode van 1998 tot 2003 11517 verschillende postcodes gevonden waar illegalen wonen die zijn aangehouden. Maandelijks kwamen er in 2003 meer dan 100 van zulke postcodegebieden bij. Op basis van gegevens van Experian, een leverancier van postcodesegmentaties, zijn deze gebieden onder meer sterk gerelateerd aan economische indicatoren als de koopkracht en de schuldpositie van de inwoners. Deze gegevens en wetmatigheden maken het mogelijk een goed gestratificeerde steekproef van postcodegebieden te trekken. Daarna volgt echter het meest onzekere gedeelte, het maken van schattingen van aantallen illegalen in een postcodegebied. Hiervoor zijn verschillende benaderingen mogelijk: (i) vaststellen door opsporingsambtenaren, (ii) waarneming door getrainde deskundigen, (iii) face to face enquêtes door goed getrainde interviewers in de postcodegebieden (iv) ondervraging via een groot internetpanel (bij TNS NIPO bleken er voldoende respondenten in de postcodegebieden met veel illegalen te wonen). Onbekend is echter of dit ook daadwerkelijk valide schattingen per postcodegebied oplevert. Een positieve aanwijzing in die richting is het feit dat in de Verenigde Staten het aantal illegalen op basis van face to face-enquêtes wordt vastgesteld, met een acceptabele onderrapportage, maar het is de vraag of deze bevinding ook van toepassing is op de Nederlandse situatie.

Randomized response

Randomized response is een methode voor het meten van sensitieve meningen, attitudes of gedrag, zoals wetsovertredingen, het drugs- en alcoholgebruik en seksualiteit. Vragen over dit soort onderwerpen worden door de respondenten vaak als bedreigend ervaren waardoor zij de neiging hebben om sociaal wenselijke antwoorden te geven. Randomized response is een methode die speciaal ontworpen is om het bedreigende karakter van de vragen te minimaliseren en zo de validiteit van de antwoorden te verhogen. In randomized response worden de antwoorden op de gevoelige vragen mede door kans bepaald. Dit kan op verschillende manieren. Zo gooit de respondent in de zogenaamde 'forced response'-methode bijvoorbeeld 2 dobbelstenen en beantwoordt vervolgens de vragen op basis van het gegooid aantal ogen. Bij 2, 3 of 4 is 'ja' verplicht en bij 11 of 12 is 'nee' verplicht. Bij alle overige worpen antwoordt de respondent naar waarheid. Op die manier weet niemand of een waar of een verplicht antwoord is gegeven en is de privacy van de respondent

gewaarborgd. Op basis van statistische analyse van de gegeven antwoorden kan de verdeling van de ware antwoorden worden berekend. Randomized response is moeilijk denkbaar bij rechtstreeks onderzoek onder potentiële illegalen (“bent u illegaal?”, ja/nee), maar het is wel een methode die gebruikt kan worden voor het verkrijgen van informatie over illegalen, bijvoorbeeld bij werkgevers die mogelijk illegalen in dienst hebben.

De Delphi-methode

De Delphi-methode is geen statistische techniek, maar een procedure die is gebaseerd op het combineren van inschattingen van deskundigen. De basisprocedures van de Delphi-methode zijn de volgende. (1) Er wordt een panel van deskundigen samengesteld die ieder een eigen kijk hebben op het probleemgebied. (2) De deskundigen komen niet direct met elkaar in contact. Ze nemen kennis van elkaars inbreng via een moderator die de input anoniem distribueert. (3) De deskundigen geven de eerste ronde hun mening over een bepaald probleemgebied, in dit geval de omvang van een verborgen populatie. (4) In de tweede ronde worden ze met elkaars mening geconfronteerd; ze stellen hun mening bij en/of motiveren hun mening in de tweede ronde. (5) Deze procedure wordt in volgende rondes herhaald, idealiter tot de meningen convergeren (en er dus één gezamenlijke schatting van de omvang van de populatie is) of anders tot de meningen niet meer veranderen. De methode is bedoeld om de gezamenlijke kennis van de deskundigen te benutten, zonder last te hebben van de problemen die in persoonlijke discussies een rol kunnen spelen, zoals conflicten en het domineren van de meest verbaal vaardige personen. Hoewel de methode op zeer grote schaal, voor een grote variëteit aan problemen wordt gebruikt, is ze niet onomstreden; er wordt in de regel geen gebruik gemaakt van een inzichtelijke empirische bewijsvoering. Toch is er wel statistisch onderzoek naar de methode gedaan. Niet verrassend, maar wel prettig, is dat er overweldigend bewijs is dat de consensus na de eerste ronde toeneemt. Kennelijk vinden de deskundigen elkaar in de regel wel. Daarnaast blijkt uit vergelijking met reeds bekende cijfers of voorspellingen die inmiddels gecontroleerd kunnen worden dat de uitkomsten van Delphi aan het eind nauwkeuriger zijn dan in de eerste ronde. Ook levert Delphi nauwkeuriger resultaten dan andere groepsprocessen. Met andere woorden: daar waar een wetenschappelijke evaluatie van Delphi mogelijk is blijkt de methode beter te zijn dan vergelijkbare alternatieven. Toch lijkt het gebruik van de Delphi-methode als basis voor een monitor problematisch. Niet statistisch geschoolden zullen dan vanuit hun lokale ervaring cijfers moeten extrapoleren naar nationaal niveau. Het zal bovendien moeilijk zijn de redeneringen die worden gebruikt van jaar tot jaar constant te houden, waarmee de vergelijkbaarheid van de cijfers over de tijd heen twijfelachtig wordt.

Overige methoden

Hoofdstuk 2 besluit met een korte beschrijving van een aantal andere methoden. Met de *demografische methode* wordt bij een gegeven startpunt in de tijd op basis van demografische variabelen als geboorte, sterfte en emigratie berekend hoe de verwachte omvang van een cohort illegalen zich in de tijd ontwikkelt. De *multiplier-methode* is een manier om bekende verhoudingen te gebruiken om een onbekend aantal te schatten. Zo kan men gebruik maken van het gegeven dat ongeveer 1 op de 10 gebruikers van heroïne vroeg of laat overlijdt aan een overdosis. Het aantal slachtoffers van een overdosis in een bepaalde tijd of plaats wordt dan gebruikt om het totaal aantal gebruikers te schatten, dus als er 50 slachtoffers van een overdosis zijn, dan schat men het aantal heroïnegebruikers op 500. Met de *residuenmethode* wordt het aantal illegalen berekend als een restgrootheid. De gebruikelijke manier om dit te doen is een schatting te maken van het totaal aantal allochtonen, en hier het aantal legale allochtonen van af te trekken. Ten slotte is de *drie kaartenmethode* een nieuw alternatief voor randomized response om gevoelige vragen te kunnen stellen. Door de vraagteksten op drie verschillende kaarten, waarvan de respondent er één krijgt, slim te variëren, kan een schatting worden gemaakt van het percentage dat in een gevoelige categorie valt.

Het gebruik van de methoden in de praktijk

Waar de opsomming van de verschillende methoden een redelijk overzichtelijk geheel van alternatieven voor het schatten van aantallen illegalen vormt, is een overzicht van met name de internationale praktijk een rommelig geheel van half gelukke pogingen, ontbrekende informatie en ad hoc aanpassingen waar weinig systeem in zit. Zo werd in *Marokko* op basis van een demografische analyse het aantal emigranten bepaald. Door hiervan het aantal legale emigranten af te trekken, kon de illegale emigratie worden afgeleid. Tot hiertoe verliep de berekening op een plausibele manier. Vervolgens werd echter dit aantal toegerekend aan de verschillende Europese landen evenredig aan de legale emigratie. Voor deze berekening was echter geen empirische basis. Een poging om een dergelijke analyse ook voor *Tunesië* te doen liep stuk op de slechte kwaliteit van de Tunesische volkstelling. In *Portugal* werden immigratiegegevens vergeleken met volkstellingsgegevens van 1981. Het aantal illegalen werd geschat als het verschil tussen het aantal buitenlanders in de volkstelling en het aantal als legaal geregistreerde immigranten. In 1991 bleek er verwarring te zijn over het verschil tussen geboorteplaats en woonplaats, waardoor de schatting achteraf waardeloos bleek. In *België* stelde men in 1990 vast dat in de buurlanden de illegale populatie

werd geschat op ongeveer 10% van de totale allochtone populatie. Dit cijfer werd overgenomen, maar nooit als officieel cijfer gepubliceerd. In *Tsjechië* werd een studie gedaan naar mazen in de wet. Vervolgens werd door deskundigen (onderzoekers, immigratie-officials) beoordeeld in welke mate illegalen en werkgevers hier gebruik van maakten. Helaas zijn hiervan de resultaten nooit bekend gemaakt.

Dit soort voorbeelden (er staan er veel meer in hoofdstuk 3) maken duidelijk dat pogingen om illegalen te tellen alleen zin hebben wanneer er serieus hulpbronnen en kwaliteit in wordt geïnvesteerd. Twee landen waar dit inderdaad is gebeurd zijn de Verenigde Staten en Zuid-Afrika. Het land waar de residuenmethode meer serieus wordt toegepast is de *Verenigde Staten*. Hier kan men gebruik maken van de volkstelling, het aanvullende ACE (Accuracy and Coverage Evaluation) en de CPS (Current Population Survey). Deze gegevens worden gecombineerd met die van de vreemdelingendienst. Het US Bureau of the Census is in 1980 begonnen met het tellen van illegale immigranten, gebaseerd op de volkstelling. Op basis van de vragen in de volkstelling werd in 1980 geschat wie allochtoon was en wie er al dan niet legaal was.

Het algemene kader om het aantal illegalen te schatten was een formule waarmee het aantal allochtonen in de Verenigde Staten werd berekend:

$$\text{Aantal allochtonen} = L - (M+E) + T + R$$

- L het aantal legale allochtonen
- M de totale sterfte onder allochtonen
- E de totale emigratie van allochtonen
- T het aantal tijdelijk verblijvende allochtonen
- R het aantal ongeautoriseerde allochtonen; dit werd gesplitst in het aantal allochtonen dat nog in een asielprocedure zit, en de echte illegalen.

De componenten werden geschat op basis van verschillende bronnen, waaraan door verschillende teams is gewerkt.

In *Zuid-Afrika* werden de schattingen gebaseerd op stroomgegevens (immigratie en emigratie) van drie groepen: (1) legaal binnengekomen, illegaal verblijf, (2) illegaal binnengekomen, legaal verblijf en (3) illegaal binnengekomen, illegaal verblijf. Daarnaast is er een groot onderzoek gedaan in de landen van herkomst naar de aard van het illegaal verblijf. Een verrassende uitkomst was dat het verblijf van illegalen in Zuid-Afrika in de regel kort was (enkele maanden), waarna men weer naar het land van herkomst terugkeerde. Dit was voor de onderzoekers reden om het illegalenprobleem in Zuid-Afrika enigszins te relativiseren.

Onderzoek in Nederland

Binnen Nederland is een rijke ervaring met het schatten van aantallen illegalen; met name op het gebied van meer geavanceerde statistische methoden is er binnen Nederland veel werk verricht. Er is een beperkt aantal publicaties waarin de nadruk ligt op illegale werknemers. Zo zijn er al publicaties uit 1994, waarin schattingen worden gepresenteerd op basis van indrukken van de Loon Technische Dienst en de Belastingdienst en op basis van interviews met werkgevers. Recent, in 2005, is in een rapport van Regioplan gepubliceerd over naleving van de Wet Arbeid Vreemdelingen. Hierin werd onder meer de *randomized response*-methode toegepast.

In 1995 is een schatting gemaakt van het aantal illegalen in de vier grote steden op basis van een schatting voor de stad Rotterdam. De schatting voor Rotterdam kwam tot stand door een schatting te maken van het aantal criminele illegalen in Rotterdam, en deze te combineren met de ratio criminele/niet-criminele illegale vreemdelingen verkregen uit eigen interviews. Deze combinatie is aan te merken als een toepassing van de *multiplier*-methode.

De methoden waarover het meest is gepubliceerd zijn de vangst-hervangstmethode. Met de beide hoofdvarianten van de methode, gekoppelde registraties en herhaald vangen is ervaring opgedaan. Zo zijn er op basis van *herhaald vangen* schattingen voor de illegale populatie over de periode 2000-2003 gemaakt op basis van het regionale Vreemdelingen Administratie Systeem (VAS), dat eind maart 2005 is opgeheven en vervangen door het landelijke Politie Suite Handhaving-Vreemdelingen (PSH-V). In principe werden alle aangehouden illegalen hierin geregistreerd. Daarbij moest het gebruikte model worden aangepast voor illegalen die afkomstig waren uit Oost-Europa (verblijven hier in de regel maar kort, dus geen gesloten populatie) en illegalen die Nederland effectief waren uitgezet (hierop was het Poissonmodel niet van toepassing). Het feit dat de methode tot relatief grote betrouwbaarheidsintervallen leidt bleek hierbij een probleem te zijn.

Een tweede relevante in Nederland gebruikte schattingsmethode is een vangst-hervangst schatting op basis van *gekoppelde registraties*. Deze is toegepast om te komen tot een schatting van het aantal Antillianen die niet staan ingeschreven in de Gemeentelijke Basis Administratie (GBA). Gezien de relatie van de Antillen met Nederland kan hier niet van "illegale Antillianen" worden gesproken. In het algemeen bestaat er echter vaak een grote overlap tussen niet ingeschreven staan in de GBA en illegaal zijn. De gegevens werden gecombineerd met het politie HerkenningsDienst Systeem (HKS). De schattingsmethode bleek vervolgens relatief eenvoudig uit te voeren, met betrouwbaarheidsintervallen van een acceptabele omvang.

Hoe nu verder?

In hoofdstuk 4 zijn de besproken methoden systematisch vergeleken op hun bruikbaarheid in de nabije toekomst. Daarbij vallen de meeste methoden af, omdat ze voordat ze kunnen worden ingezet een aanzienlijke hoeveelheid vooronderzoek vergen, waarbij het antwoord op de vraag of ze daadwerkelijk bruikbaar zijn onzeker is. Overigens is op basis van dit rapport wel duidelijk hoe zulk vooronderzoek zou moeten worden uitgevoerd. Vangst-hervangstmethoden zijn wel direct bruikbaar; ze hebben zichzelf bewezen, al valt er op de mate waarin aan de vooronderstellingen is voldaan nog wel wat af te dingen. De methode is goedkoop, want de gegevens zijn beschikbaar. Echter, bij de methode van herhaald vangen is de omvang van de betrouwbaarheidsintervallen een probleem, bij de gekoppelde registraties kan het moeilijk worden het onderzoek op het voor het beleid voldoende recente gegevens te baseren.