

Actualisatieonderzoek schatting aantal in Nederland verblijvende Antilliaanse Nederlanders die niet ingeschreven zijn in de GBA

**Een ‘capture-recapture’-analyse
in opdracht van het Ministerie van VROM**

Peter van der Heijden, Dave Hessen en Maarten Cruyff
30 november 2009
IOPS-Utrecht
Universiteit Utrecht

Wij bedanken de heer drs. J. Oudhof (CBS) voor zijn adviezen inzake de koppeling van bestanden. Wij merken op dat hij op generlei wijze verantwoordelijk is voor de inhoud van dit manuscript. Wij danken de heren drs. L. Engbers (CBS) en ing. M.M.E.M. Schillings (CBS) voor de assistentie bij het werkend krijgen van onze programmatuur in de CBS werkomgeving.

Actualisatieonderzoek schatting aantal in Nederland verblijvende Antilliaanse Nederlanders die niet ingeschreven zijn in de GBA

**Een ‘capture-recapture’-analyse
in opdracht van het Ministerie van VROM**

Peter G.M. van der Heijden, Dave Hessen en Maarten Cruyff

Samenvatting

In 2006 is een rapportage verschenen met schattingen van het aantal Antillianen dat in Nederland verblijft zonder ingeschreven te zijn in de Gemeentelijke basisadministratie persoonsgegevens (GBA). Deze schattingen liepen van ongeveer 4 % in het jaar 2000 tot bijna 13 % in het jaar 2003. In de voorliggende rapportage worden schattingen gepresenteerd op basis van een verbeterde methode voor de jaren 2004 tot en met 2007. Uit aanvullende analyses in het voorliggende onderzoek is namelijk duidelijk geworden dat de in 2006 gepresenteerde schattingen ten dele gebaseerd waren op aannames die onjuist blijken. De schattingen op basis van de verbeterde methode over de periode 2004 – 2007 lopen van 3.0 % in 2004 tot 1.3 % in 2007. Net als in de rapportage van 2006 zijn de schattingen gebaseerd op de zgn. ‘capture-recapture’ methode. Twee bestanden zijn door het CBS gekoppeld, namelijk de GBA en het Herkenningssysteem van de politie (HKS). Door deze koppeling zijn drie groepen Antillianen te onderscheiden: (i) de groep die zowel in de GBA als in HKS bekend is; (ii) de groep die wel in HKS maar niet in de GBA bekend is; (iii) de groep die wel in de GBA maar niet in HKS bekend is. Een groep die noch in de GBA noch in HKS bekend is, dient te worden geschat. De schatting van de omvang van deze laatste groep vindt plaats met de capture-recapture methode. Door de omvang van groepen (ii) en de geschatte groep op te tellen ontstaan schattingen van het aantal Antillianen dat in Nederland verblijft zonder ingeschreven te zijn in de GBA. Er zijn uitsplitsingen van de schattingen naar leeftijd en geslacht. Er worden ook schattingen gepresenteerd van welk percentage van elke groep bekend is in HKS.

De capture-recapture methode maakt gebruik van een aantal vooronderstellingen. De houdbaarheid van deze vooronderstellingen wordt gedetailleerd besproken. In het bijzonder wordt diep ingegaan op een wijziging van de data die worden geanalyseerd: waar voor de periode 2000 – 2003 eerder alle aanhoudingen in HKS werden gebruikt, worden nu de aanhoudingen door de Koninklijke Marechaussee aan de grens (w.o. Schiphol) uitgesloten.

1. Inleiding

In 2006 is in opdracht van het Ministerie van Justitie, door het WODC een ‘capture-recapture’-analyse uitgebracht met de titel “Schatting van aantal in Nederland verblijvende Antillianen die niet ingeschreven zijn in de GBA” (zie Van der Heijden, Zwane en Hessen, 2006)¹. Er was hier sprake van een ‘quick scan’. Voorzover ons bekend is dit het enige onderzoek dat de omvang van het aantal in Nederland verblijvende Antillianen dat niet ingeschreven is in de GBA met een ‘capture-recapture’-analyse heeft geschat. In 2003 is de omvang geschat op ongeveer 11.800, destijds 13% van de totale omvang Antilliaanse Nederlanders in Nederland. Dit was een geschatte toename van 8,9 procentpunt ten opzichte van het jaar 2000, waarvoor 4,1 % werd geschat.

Recente vervolgcijfers ontbraken tot nu toe. De minister voor Wonen, Wijken en Integratie achtte recente cijfers echter van belang waarbij een vergelijkbaarheid van de methode cruciaal is voor de vergelijkbaarheid van de uitkomsten en dus voor de trendontwikkeling. Gelet daarop heeft het ministerie van VROM in 2009 opdracht gegeven een actualisatieonderzoek uit te voeren. Het resultaat hiervan ligt voor u. Bij dit herhalingsonderzoek wordt de kwantitatieve analyse vanuit zowel de omvang als de samenstelling benaderd voor de jaren 2004-2007. In aanvulling op de rapportage uit 2006 wordt in het voor u liggende onderzoek ook gekeken naar de politieregio waar Antilliaanse Nederlanders zijn aangehouden.

Inschrijving in de GBA is van belang omdat dit bij kan dragen aan het krijgen van zicht op een groep. Hoewel de GBA als algemene basisregistratie niet bedoeld is voor en derhalve ook niet ingericht is op het registreren van allerlei doel- of risicogroepen, kan de GBA door verstrekking van algemene gegevens overheidsinstanties helpen om door hen gedefinieerde risicogroepen beter in beeld te brengen. Hoe vollediger de registratie van in Nederland verblijvende personen in de GBA is, hoe meer ondersteuning de GBA hierin kan bieden. Registratie in de GBA kan zo een hulpmiddel zijn om een beter zicht op een groep te krijgen en de niet geregistreerde groep zo klein mogelijk te houden. Verantwoordelijk voor het zich inschrijven is in eerste instantie de burger zelf.

Onderzoeksvraag

Gevraagd is de omvang en samenstelling van de groep in Nederland verblijvende Antillianen die niet in de GBA staan ingeschreven in kaart te brengen en daarover te rapporteren.

Doel van het onderzoek

Het doel van de actualisatie is, om door kwantitatieve analyse een beeld te krijgen van de omvang en de samenstelling van de groep Antilliaanse Nederlanders (jongeren en ouderen) die langere tijd in Nederland verblijven, zonder in de GBA te staan geregistreerd.

De omvang

- Het is de bedoeling te weten of de omvang van de groep is af- of toegenomen (trendontwikkeling).

De samenstelling

- Het is de bedoeling ten minste te weten wat de (demografische) kenmerken van de personen uit deze groep zijn.
- De analyse moet een uitspraak kunnen doen over welke conclusies er op basis van deze analyse kunnen worden getrokken.

Onderzoeksopzet/ onderzoeksmethode/ gevraagde analyses

Verzocht is om een capture-recapture analyse voor het onderzoek, tenzij kon worden aangegeven dat een andere methode beter, zuiverder is. Desalniettemin, moest er wel een vergelijking kunnen plaatsvinden met de uitkomsten uit het eerste onderzoek. De vergelijkbaarheid van de methode is cruciaal voor de vergelijkbaarheid van de uitkomsten en dus voor de trendontwikkeling.

¹ Schatting van aantal in Nederland verblijvende Antillianen die niet ingeschreven zijn in de GBA, TK 2005-2006, 26 283, nr 29. Het onderzoek is in 2005 uitgevoerd door de Interuniversitaire Onderzoeksschool voor Psychometrie en Sociometrie (IOPS), Prof. Van der Heijden, Eugene Zwane en Dave Hessen.

Net als in de rapportage van 2006 wordt gebruik gemaakt van een koppeling van de GBA en het HerKenningsSysteem (HKS) van de politie. In aanvulling op de rapportage uit 2006 wordt in het voor u liggende onderzoek ook gekeken naar de politieregio waar Antilliaanse Nederlanders zijn aangehouden.

Het onderzoek is puur kwantitatief. Motieven om zich niet te registreren zijn niet vastgesteld.

2. Methode

De schatting heeft net als in de eerdere rapportage (van der Heijden, Zwane en Hessen, 2006) plaatsgevonden met moderne varianten van de zgn. capture-recapture methode. In deze schattingsmethode wordt gebruik gemaakt van gekoppelde gegevens van registers die aanwezig zijn op het CBS. Hiertoe worden twee bestanden gekoppeld:

- GBA gegevens
- HKS gegevens.

Na de koppeling is de volgende tabel te maken:

	Niet in GBA	In GBA
Niet in HKS	(x)	(1)
In HKS	(2)	(3)

In (3) zit het aantal mensen dat zowel in de GBA als in HKS bekend is;

In (2) zit het aantal mensen dat wel in HKS maar niet in de GBA bekend is;

In (1) zit het aantal mensen dat wel in de GBA maar niet in HKS bekend is;

Er zijn geen gegevens over de groep die noch in de GBA noch in HKS bekend is. Deze groep is aangegeven met (x).

Het aantal mensen in (x) wordt door ons geschat. De optelling van deze schatting bij het aantal in (2) levert dan het antwoord op de onderzoeksvraag: hoeveel Antillianen zijn er die in Nederland verblijven maar niet in de GBA zijn ingeschreven?

Bij de schatting zal gebruik gemaakt worden van kenmerken van de Antillianen zoals die bekend zijn in HKS en in de GBA.

Voor zowel HKS als de GBA zijn dit:

- Leeftijd
- Geslacht

Voor alleen HKS zijn dit:

- Som antecedenten
- Harddruggebruiker (ja/nee)

Voor alleen de GBA zijn dit:

- Verblijfsduur
- Burgerlijke staat.

Voor meer informatie over deze variabelen verwijzen wij naar Appendix 1.²

De door ons gebruikte statistische methoden staan beschreven in Appendix 3, Chao et al. (2001), en Zwane en Van der Heijden (2007), Van der Heijden, Zwane en Hessen (2009) en Sikkels, Van der Heijden en Van Gils (2006).

Nadere precisering van het onderzoek

Wij maken de volgende opmerkingen over de uit te voeren schattingen:

- Omdat in HKS geen onderscheid gemaakt kan worden tussen Antilliaanse en Arubaanse Nederlanders (hier gebruikt HKS dezelfde code voor), heeft het voorliggende rapport betrekking op Antilliaanse en Arubaanse Nederlanders samen, en wordt geen onderscheid tussen deze twee groepen gemaakt. Voor de leesbaarheid wordt in het vervolg de term 'Antillianen' gebruikt voor beide groepen samen.
- Omdat in HKS niet geregistreerd is of een persoon Antilliaan uit de tweede generatie is, is het niet mogelijk om bij dit onderzoek een onderscheid aan te brengen tussen verschillende generaties waaruit

² Wij merken hier op dat, omdat het aantal niet-geregistreerde Antillianen veel lager bleek te zijn te zijn dan verwacht (waarover later meer), er afgezien is van analyses waarin de variabelen betrokken zijn die alleen aanwezig zijn in HKS (dus Som antecedenten en Harddruggebruiker) en in GBA (dus Verblijfsduur en Burgerlijke staat).

Antillianen afkomstig zijn. De uitkomsten in dit onderzoek hebben slechts betrekking op eerste generatie Antillianen.³

- Het is mogelijk dat een Antiliaan zich inschrijft bij een gemeente en hierbij een briefadres opgeeft als adres.⁴ Wij schatten hier het aantal personen dat niet bekend is in de GBA, en dus noch een woonadres, noch een briefadres heeft opgegeven.
- Door de afwezigheid van de leeftijdscategorie jonger dan 12 jaar binnen HKS, is er voor deze leeftijdscategorie geen schatting te maken met standaardmethoden. Onze schattingen hebben daarom alleen betrekking op personen van 12 jaar en ouder.

Aannamen van schattingen

Bij het maken van de schatting is een aantal gebruikelijke aannamen gemaakt. Het effect van mogelijke schendingen van deze aannamen zullen we hier bespreken.

Een eerste aanname is dat er sprake is van de omvang van *een populatie die gesloten is*. Hierbij kan men denken aan een schatting van het aantal vissen in een aquarium: deze populatie is gesloten indien, tijdens het verzamelen van de data, er geen vissen worden geboren (toename van populatie) of sterven (afname van populatie). Deze aanname is hier geschonden, omdat er tijdens ‘het verzamelen van de data’ (hier: de periode waarop de registratie betrekking heeft) Antillianen Nederland zullen binnen komen en zullen verlaten. Dit probleem is door ons verkleind door schattingen per jaar te maken, in plaats van over een periode van meerdere jaren. Probleem blijft dat personen die niet gedurende een geheel jaar aanwezig zijn geweest, in principe een kleinere kans hebben om waargenomen te worden in de GBA en binnen HKS. Wij kunnen niet goed inschatten hoe groot de invloed van de schending van deze aanname is op de schatting van het aantal Antillianen dat zich niet heeft ingeschreven in de GBA. De schending is met name groot indien de groep Antillianen die in HKS worden geobserveerd maar niet in de GBA, voor een groot deel bestaat uit personen waar we niet in geïnteresseerd zijn, zoals vakantiegangers. Deze schending zal er toe leiden dat de schattingen die wij rapporteren, te groot zijn.

Een tweede aanname is dat de koppeling van de bestanden zonder fouten plaatsvindt. Er zijn twee soorten fouten. De eerste is dat een record in HKS wordt gekoppeld aan een record in de GBA, terwijl het hier niet dezelfde persoon betreft. Door dergelijke fouten wordt het aantal personen dat zowel in het ene als in het andere bestand wordt aangetroffen, abusievelijk te hoog (de frequentie in cel (4) van de tabel hierboven is 1 te hoog, en die van (2) en (3) zijn 1 te laag). De tweede is dat een record in HKS *niet* wordt gekoppeld aan een record in de GBA, terwijl het hier wel dezelfde persoon betreft. Door dergelijke fouten wordt het aantal personen dat zowel in het ene als in het andere bestand wordt aangetroffen, abusievelijk te laag (de frequentie in cel (4) van de tabel hierboven is 1 te laag, en die van (2) en (3) zijn 1 te hoog). De eerste fout leidt tot een te lage schatting van de frequentie in cel (1), en de tweede fout leidt tot een te hoge schatting van de frequentie in cel (1). Informatie over de koppeling wordt gegeven in CBS (2005). In Appendix 2 geven wij meer informatie over de kans op de twee soorten fouten die wij zojuist hebben besproken. Wij zijn geneigd aan te nemen dat de koppelingsfouten gering zijn.

Een derde aanname is dat de kans om waargenomen te worden binnen HKS en de kans om waargenomen te worden binnen de GBA, onafhankelijk van elkaar zijn. Door in de eerste analyse te controleren voor de variabelen leeftijd en geslacht, en in de tweede analyse te controleren voor de variabelen leeftijd, geslacht, burgerlijke staat, verblijfsduur, antecedenten en het al dan niet zijn van een harddruggebruiker, is de kans op schending van deze aanname door ons verkleind, maar nog wel aanwezig.

- Indien deze aanname geschonden is doordat er een *positieve* samenhang tussen deze kansen is, dat wil zeggen dat personen die zich *vaker* inschrijven bij de GBA ook *vaker* in HKS voorkomen, dan dienen de door ons gepresenteerde schattingen geïnterpreteerd te worden als *onderschattingen* van de werkelijke populatie: de werkelijke populatieomvang is dan groter dan wat wij hebben geschat.

³ Inhoudelijk zou het ook niet zo voor de hand liggen om onderzoek te doen naar niet-geregistreerde tweede generatie Antillianen die door hun geboorte in Nederland hun leven al zijn begonnen als in Nederland geregistreerde (en zich dus eerst zouden moeten uitschrijven). Dit aantal is naar alle waarschijnlijkheid erg laag.

⁴ Dit moet overigens wel een bestaand woonadres zijn (en kan dus niet een postbus zijn).

- Indien deze aanname geschonden is doordat er een *negatieve* samenhang tussen deze kansen is, dat wil zeggen dat personen die zich *vaker* inschrijven bij de GBA juist *minder vaak* in HKS voorkomen, dan dienen de door ons gepresenteerde schattingen geïnterpreteerd te worden als *overschattingen* van de werkelijke populatie: de werkelijke populatieomvang is dan *kleiner* dan wat wij hebben geschat.

Voorzover deze schendingen voorkomen komt het ons voor dat de kans op een overschatting aannemelijker is dan de kans op een onderschatting: wij stellen ons dan rondzwervende jongeren (geen registratie in de GBA) voor die in aanraking komen met de politie (geregistreerd in HKS). Indien deze schending plaatsvindt, weten wij nog niet de sterkte van de schending, en wij weten dus niet hoever de werkelijke populatieomvang afligt van de door ons gepresenteerde schatting. Wij weten echter wel de bovengrens, dus als we voor een bepaalde groep een schatting van 1000 hebben, dan weten we dat het er in ieder geval niet meer zijn.

Een vierde aanname is dat de kans om voor te komen in HKS en in de GBA voor iedereen identiek is. De schending van deze aanname is beperkt door te controleren voor de variabelen geslacht en leeftijd in de eerste analyse, en een groter aantal in de tweede analyse. Hierdoor is de aanname dat de kans om voor te komen identiek is voor eenieder met dezelfde waarden op deze variabelen. In de 'quick- scan' rapportage uitgebracht in 2006 werd voorzichtig geconcludeerd dat de invloed van deze mogelijke schending waarschijnlijk gering zou zijn. Nadere analyse van de gegevens uit 2004 tot 2007 geeft echter aan dat dit voor deze jaren een onjuiste veronderstelling is: de invloed van deze schending blijkt zeer groot te zijn. In paragraaf 3.1 wordt hier nader op ingegaan.

Dit brengt ons tot de volgende conclusies.

- Indien aannamen 1 en 3 zijn geschonden, heeft dit geleid tot te hoge schattingen. Voor aanname 1 en 3 is dit door ons niet te verifiëren.
- Voor aanname 4 hebben we door nadere analyse van de nieuwe data kunnen verifiëren dat deze aanname ernstig geschonden is in de periode van 2004 – 2007 (nieuwe data), en gezien de aard van de schending ook in de voorliggende jaren 2000 – 2003 (gerapporteerd in de quick scan van 2006).

Om voor de schending van aanname 4 te corrigeren, zijn – naast analyses op de gegevens van 2004 – 2007 die voor wat betreft de aanpak identiek zijn aan die uitgevoerd voor de periode 2000 – 2003 – ook analyses uitgevoerd op aangepaste gegevens voor de periode 2004 – 2007. Voor deze laatste analyses geldt dat er geen aanwijzingen zijn dat voor deze aangepaste gegevens aanname 4 ernstig geschonden is.

3. Resultaten

3.1 Beschrijving van de te analyseren data

Hieronder vindt u de aantallen op basis van de GBA en HKS. Voor de periode 2000 – 2003 waren deze al te vinden in het ‘quick scan’ rapport dat gepubliceerd is in 2006. Voor de periode 2004 – 2007 is er voor de HKS gegevens een onderscheid gemaakt in aanhoudingen door de Koninklijke Marechaussee (KMAR) en aanhoudingen door politie in een van de politieregio’s.

2000	Niet in GBA	In GBA
Niet in HKS	?	64.247
In HKS	201	5.095

2001	Niet in GBA	In GBA
Niet in HKS	?	69.337
In HKS	428	5.650

2002	Niet in GBA	In GBA
Niet in HKS	?	71.468
In HKS	786	6.647

2003	Niet in GBA	In GBA
Niet in HKS	?	72.100
In HKS	949	6.940

2004	Niet in GBA	In GBA
Niet in HKS	?	70.522
In HKS, geen KMAR	190	6.890
In HKS, KMAR	552	322

2005	Niet in GBA	In GBA
Niet in HKS	?	69.168
In HKS, geen KMAR	131	6.686
In HKS, KMAR	136	145

2006	Niet in GBA	In GBA
Niet in HKS	?	69.013
In HKS, geen KMAR	67	6.407
In HKS, KMAR	51	93

2007	Niet in GBA	In GBA
Niet in HKS	?	69.772
In HKS, geen KMAR	81	6.330
In HKS, KMAR	52	76

Enkele zaken vallen hier op:

- Beschouwen we het totaal in de kolom “In GBA”, dan zien we een groei van het aantal in de GBA ingeschrevenen van ongeveer 69.000 in 2000 (namelijk $64.247 + 5.095$) tot 79.000 in 2003. Vervolgens vindt er tot 2007 een daling plaats tot ongeveer 76.000.
- Vervolgens beschouwen we het totaal in de rij(en) “In HKS” (in de jaren 2000 – 2003 is er slechts een enkele rij en in 2004 – 2007 zijn er twee rijen). Er is in de periode 2000-2003 een sterke toename van het aantal personen dat in HKS geregistreerd staat, namelijk van ongeveer 5.200 in 2000, tot 7.900 in 2003, en in de periode 2004 – 2007 daalt dit weer van ongeveer 8000 tot 6.500.
- Bekijken we het aantal personen dat *niet* in de GBA bekend is maar wel in HKS is geregistreerd, dan zien we in de periode 2000 – 2003 een stijging van ongeveer 200 tot 950, en in 2004 – 2007 een daling van ongeveer 740 tot 125.

Tot zover de vergelijking van de gegevens die zowel voor de periode 2000 – 2003 als 2004 – 2007 beschikbaar zijn. Nu spitsen wij ons toe op de gegevens voor de periode 2004 – 2007, die uitgesplitst zijn naar ‘aangehouden, door de Koninklijke Marechaussee’ (KMAR) en ‘aangehouden, niet door de Koninklijke Marechaussee’ (niet-KMAR). Hier vallen de volgende zaken op:

- De KMAR aanhoudingen van personen die *niet* bekend zijn in de GBA dalen van ongeveer 550 in 2004 tot ongeveer 50 in 2007.
- Ook de niet-KMAR aanhoudingen van personen die *niet* bekend zijn in de GBA dalen, maar minder scherp, van 190 tot 81.
- Beschouwen we *alle* HKS-geregistreerden, en splitsen we die uit in niet-KMAR versus KMAR, en in Niet in GBA versus GBA, dan vinden we de 2x2 tabel onderaan de tabellen voor 2004 – 2007. In deze tabellen is een zeer sterke samenhang: KMAR-aanhoudingen zijn vaak van personen die *niet*-in de GBA bekend zijn (bijvoorbeeld in 2004 552 van de $(552 + 322 =) 874$ KMAR aanhoudingen zijn van personen die niet in de GBA bekend zijn), en *geen*-KMAR aanhoudingen zijn vooral aanhoudingen van personen die *wel* in de GBA bekend zijn (190 van de $(190 + 6.890 =) 7.080$).

De in het laatste aandachtsstreepje geconstateerde samenhang is van groot belang: indien we deze samenhang zouden negeren door voor 2004 – 2007 niet-KMAR op te tellen bij KMAR, dan wordt de vierde aanname van de ‘capture-recapture’-analyse geschonden. De vierde aanname is dat voor eenieder de kans om niet in de GBA te worden aangetroffen identiek is, maar nu blijkt dat die kans voor personen aangehouden door de Koninklijke Marechaussee veel groter te zijn.

Alvorens een oplossing voor dit probleem voor te stellen, bekijken we de oorzaak van deze samenhang. De Koninklijke Marechaussee is vooral werkzaam bij de grenzen van Nederland. Aanhoudingen van Antillianen door de Koninklijke Marechaussee vinden vooral plaats op Schiphol. Waarschijnlijk gaat het hier om het vervoeren van drugs (‘bolletjes’). In de periode van 2000 tot 2003 nam dit verschijnsel en het aantal aanhoudingen toe, maar door verscherpte controle, ook in de Antillen, is dit verschijnsel sindsdien ernstig afgenomen. Dit verklaart de afname van het aantal aanhoudingen door de Koninklijke Marechaussee na 2003. Indien de gedachte juist is, dat het bij de KMAR aanhoudingen vooral drugszaken betreft, dan kunnen we vervolgens concluderen, op basis van de tabellen 2004 – 2007, dat deze drugs-aanhoudingen vooral Antillianen betrof die niet bekend waren in de GBA.

Dit probleem pakken we als volgt aan.

- Voor de periode 2004 – 2007 tellen we niet de rijen ‘niet-KMAR’ en ‘KMAR’ op, omdat hiermee aanname 4 ernstig is geschonden.
- Een analyse gebaseerd op louter de KMAR-aanhoudingen wordt ook verworpen, omdat het aannemelijk is dat er hier sprake is van een sterke negatieve relatie tussen ingeschreven zijn in HKS en ingeschreven zijn in de GBA: als men aangehouden wordt door de KMAR is men juist *niet* ingeschreven in de GBA. Dit is een schending van aanname 3, die stelt dat aangetroffen worden in GBA en aangehouden worden door de Koninklijke Marechaussee statistisch onafhankelijk zou moeten zijn. Indien er sprake is van drugstoerisme, dan wordt met de KMAR gegevens ook de aanname 1 geschonden, want dan is er sprake van een open populatie.

- Een analyse gebaseerd op de aanhoudingen die hebben plaatsgehad in politieregio's lijkt een redelijke optie. We schatten dan het aantal Antillianen dat niet ingeschreven staat in de GBA *en aan te treffen is in politieregio's*.

We concluderen het volgende.

- We gaan schattingen uitvoeren voor de periode 2004 – 2007 op basis van de aanhoudingen in politieregio's (niet-KMAR). Deze worden gerapporteerd in de volgende paragraaf.
- Volledigheidshalve voeren we voor de periode 2004 – 2007 ook de schattingen uit volgens de methodiek van de 'quick scan' rapportage. Zo is er een tijdreeks voor de gehele periode 2000 – 2007. Omdat voor deze tabellen geldt dat aanname 3 is geschonden (en ook de trend in deze tijdreeks gewantouwd dient te worden aangezien de schending niet constant is over de jaren heen), vormen deze tabellen geen onderdeel van dit rapport. In plaats daarvan worden zij separaat aan de opdrachtgever aangeboden.
- Van groot belang is te constateren dat de resultaten die eerder in de 'quick scan' van 2006 zijn gerapporteerd, gebaseerd blijken te zijn op een aanname die geschonden is. De KMAR aanhoudingen hadden voorafgaand aan de analyses uit de data verwijderd dienen te worden. Met de 'quick scan' is daardoor een onjuist beeld geschetst van het aantal Antillianen dat in Nederland verblijft zonder ingeschreven te zijn in de GBA. De invloed van de schending is namelijk groot. Bezien we bijvoorbeeld het jaar 2000, dan betekent de aanname van onafhankelijkheid tussen opname in de GBA en opname in HKS (die aan de basis van de schattingsmethode ligt) dat, $\frac{?}{201} = 64.247/5.095$, zodat $? = 201 * 64.247/5.095$. Omdat dit aantal stijgt van 201 (in 2000) tot 949 (in 2003), stijgt de schatting van het vraagteken ook met ongeveer een factor 5. Wij hebben zojuist geconcludeerd dat ook in het aantal van 949 Antillianen geregistreerd in HKS maar niet in de GBA, zonder enige twijfel veel Antillianen zitten die aangehouden zijn door de Koninklijke Marechaussee. Het ware beter geweest deze personen niet mee te nemen in de analyse. We concluderen dat, indien we deze KMAR aanhoudingen uit de data hadden verwijderd, de schattingen van het aantal Antillianen dat niet staat ingeschreven in de GBA stond voor bijvoorbeeld 2003 veel lager was uitvallen.

3.2 Schattingen gecontroleerd voor leeftijd en geslacht

In een eerste reeks analyses hebben wij onderzocht of Leeftijd en Geslacht van invloed zijn op de kans niet in de GBA te worden aangetroffen.

De eerste analyseresultaten zijn die voor het model waarin de kans om in de GBA en in HKS voor te komen afhangt van Geslacht en Leeftijd samen. Het gaat hier om het loglineaire model (Geslacht x Leeftijd x HKS), (Geslacht x Leeftijd x GBA). Anders gezegd, het model neemt aan dat HKS en GBA onafhankelijk zijn in iedere combinatie van Geslacht en Leeftijd.

Voor 2007 treft u in onderstaande tabel de resultaten aan. Voor gedetailleerde gegevens over 2004 – 2007 verwijzen we naar Bijlage 1. In Bijlage 1 zijn ook betrouwbaarheidsintervallen van geschatte aantallen te vinden.

2007		Niet in GBA	in GBA	Totaal	% niet in GBA	% in HKS
<i>Man</i>	<i>13-17</i>	11.3	2133	2144.3	0.5	17.8
<i>Man</i>	<i>18-24</i>	117.1	5830	5947.1	2.0	19.1
<i>Man</i>	<i>25-44</i>	234.3	16914	17148.3	1.4	16.0
<i>Man</i>	<i>45+</i>	26.8	12523	12549.8	0.2	7.5
<i>Vrouw</i>	<i>13-17</i>	35.3	2066	2101.3	1.7	5.8
<i>Vrouw</i>	<i>18-24</i>	93.7	5868	5961.7	1.6	5.4
<i>Vrouw</i>	<i>25-44</i>	174.1	15996	16170.1	1.1	4.1
<i>Vrouw</i>	<i>45+</i>	328.3	14772	15100.3	2.2	1.2
Totaal		1021.0	76102	77123.0	1.3	8.5

We leggen hier de eerste regel uit:

- ‘Niet in GBA’ wil zeggen: het aantal personen dat in HKS zit maar niet in de GBA (dus in cel (2)) plus het geschatte aantal personen dat noch in HKS noch in de GBA zit (cel (x)). Voor jongens van 13-17 zijn dit er in 2007 dus 11.
- ‘in GBA’ wil zeggen: het aantal personen dat in de GBA zit, al dan niet in HKS (dus cel (1) en (3) samen). Voor jongens van 13-17 zijn dit er in 2007 dus 2.133.
- ‘Totaal’ wil zeggen: de som van ‘Niet in GBA’ en ‘in GBA’ dus het totale aantal Antillianen dat in Nederland verblijft. Voor jongens van 13-17 zijn dit er in 2007 dus 2144.
- ‘% niet in GBA’ wil zeggen: het eerste getal gedeeld door het totaal. Voor jongens van 13-17 zijn dit er in 2007 dus $100 * 11.3 / 2133 = 0.5$ %.
- ‘% in HKS’: is berekend door het aantal personen, geregistreerd én geschat, dat in HKS staat te delen door het eerder besproken Totaal. Voor jongens van 13-17 komt in 2007 dus 17,8 % voor in HKS.

Wij noemen hier enkele zaken die ons in de tabellen van 2004 – 2007 als beleidsmatig interessant voorkomen:

- het percentage niet in GBA geregistreerde personen daalt van 3.0 % in 2004 tot 1.3 % in 2007. Het betreft hier in 2007 ongeveer 1000 personen (met een 95 % betrouwbaarheidsinterval van 667 – 1400).
- in alle leeftijdscategorieën komt het niet registreren voor, en zowel bij mannen als vrouwen. In 2007 lopen de percentages van 0.2 % tot 2.2 %, waarbij benadrukt moet worden dat de betrouwbaarheidsintervallen voor deelgroepen groter zijn (en dus minder betrouwbaar) dan voor het totaal.
- Hoewel dit niet tot de onderzoeksvraag behoort, hebben wij ook percentages gegeven van de Antillianen in Nederland die in HKS voorkomen. Deze zijn lager bij de vrouwen (ongeveer 5.5 % bij jonge vrouwen) en hoger bij de jongere mannen (ongeveer 18 %). Cijfers hebben betrekking op alleen de eerste generatie. Wij merken op dat deze percentages identiek zijn onder de Antillianen die geregistreerd zijn in de GBA en zij die dit niet zijn, omdat dit een van de aannamen van onze analyses is.

We merken tenslotte met nadruk op dat de hier gerapporteerde schattingen veel lager uitvallen dan die uit de 'quick scan' van 2006. Waar hier het percentage in 2004 op 3.0 % ligt, lag dit in 2003 in de 'quick scan' op 13.0 %. Dit laatste percentage geeft een verkeerd beeld omdat het gebaseerd is op een aanname die geschonden is (zie sectie 3.1).

4. Conclusie

Het onderzoek kende de volgende doelen:

De omvang

- Het is de bedoeling te weten of de omvang van de groep is af- of toegenomen (trendontwikkeling).

De samenstelling

- Het is de bedoeling ten minste te weten wat de (demografische) kenmerken van de personen uit deze groep zijn.
- De analyse moet een uitspraak kunnen doen over welke conclusies er op basis van deze analyse kunnen worden getrokken.

V.w.b. de *omvang* kan het volgende worden gesteld.

- in het voorliggende onderzoek is gebleken dat een van de aannames voor de analyses gerapporteerd in de 'quick scan' van 2006 geschonden is. Daardoor zijn de schattingen gerapporteerd in 2006 over de periode 2000 – 2003, nl. van ongeveer 4 % oplopend naar 13 %, te hoog geweest. Het achteraf corrigeren van deze schattingen maakte geen deel uit van onderhavig onderzoek.

In onderhavig onderzoek is de methode voor de schattingen aangepast naar aanleiding van de gebleken schending. De verbeterde schattingen voor de periode 2004 – 2007 laten voor 2004 een schatting van 3.0 % zien, die afneemt naar 1.3 % voor het jaar 2007. Deze percentages refereren naar de populatie Antillianen die in HKS opgenomen waren ten gevolge van een aanhouding in een politieregio, en deze percentages refereren *niet* naar Antillianen die aangehouden zijn aan de grens.

V.w.b. de *samenstelling* kan het volgende worden gesteld. Doordat het aantal niet-geregistreerde Antillianen in 2006 en 2007 teruggelopen is tot ongeveer 1.000, is een demografische beschrijving moeilijk te maken omdat de schatting van 1.000 personen gebaseerd is op de kenmerken van ongeveer 75 personen, en alle schattingen van opsplitsingen van deze groep (bijvoorbeeld in geslacht en leeftijd) daarom grote betrouwbaarheidsintervallen kent. Deze betrouwbaarheidsintervallen zijn evenwel gerapporteerd in Bijlage 1 van dit rapport.

Beperkingen

Net als het onderzoek gepubliceerd in 2006, kent ook dit onderzoek een aantal beperkingen:

- Het is gezien de wijze van registreren in de HKS niet mogelijk om een onderscheid aan te brengen tussen Antillianen en Arubanen.
- De uitkomsten in dit onderzoek hebben slechts betrekking op eerste generatie Antillianen.
- Het is mogelijk dat een Antiliaan zich inschrijft bij een gemeente en hierbij een postadres opgeeft als adres. Wij schatten hier het aantal personen dat niet bekend is in de GBA, en dus noch een woonadres, noch een postadres heeft opgegeven⁵.
- Door de afwezigheid van de leeftijdscategorie jonger dan 12 jaar binnen HKS, is er voor deze leeftijdscategorie geen schatting te maken met standaardmethoden. Onze schattingen hebben daarom alleen betrekking op personen van 12 jaar en ouder.

Er zijn bij het maken van de schattingen aannamen gemaakt die uitgebreid beschreven zijn in paragraaf 2 en 3.1. Een mogelijke schending is dat onder de personen die wel in HKS voorkomen maar niet in de GBA, veel kort in Nederland verblijvenden zitten (bijvoorbeeld vakantiegangers, open populatieprobleem). In dat geval is het aantal geschatte personen te hoog. Een andere mogelijke schending is dat er een relatie is tussen het wel ingeschreven zijn in HKS maar niet in de GBA. Indien dit het geval is, dan leidt dit ook tot te hoge schattingen. Wij concluderen dat, indien deze aannamen zijn geschonden (wat overigens niet bekend is), dit waarschijnlijk heeft geleid tot te hoge schattingen.

⁵ Het aantal Antillianen dat een postadres opgeeft, is ons onbekend, omdat in GBA een postadres niet van een gewoon adres is te onderscheiden.

Het onderzoek is puur kwantitatief. Motieven om zich niet te registreren zijn niet vastgesteld.

Appendix 1: Gebruikte variabelen

Het door het CBS samengestelde bestand staat bekend onder de naam ‘GBA-selectiebestand’, het bestand aangeleverd door de KLPD onder HKS. Wij hebben de volgende variabelen (aangegeven met hun variabele naam) gebruikt uit beide bestanden:

	GBA-selectiebestand	HKS
Leeftijd	<i>GEBDATJ</i>	<i>lft_ldel</i>
Geslacht	<i>GESL</i>	<i>gesl_cat</i>
Regio		<i>woon_reg</i>
Verblijfsduur	<i>BEGGELDJ</i>	
Burgerlijke staat	<i>BURGST</i>	
Som antecedenten		<i>somant2</i>
Harddruggebruiker		<i>versl_cat</i>

Geslacht heeft de indeling (i) man (ii) vrouw

Leeftijd: hier is gekozen voor de indeling (i) tot 12 jaar (ii) 13-17 (iii) 18-24 (iv) 25-44 en (v) 45-rest. De leeftijd tot 12 jaar wordt alleen geconstateerd binnen de GBA en niet binnen HKS, en daarom is met standaard methoden geen schatting te maken van het aantal kinderen dat niet binnen de GBA voorkomt.

Regio: hier is gekozen voor een politieregio-indeling, en wel als volgt: (i) de 25 Nederlandse politieregio's (ii) regio onbekend. De laatste categorie komt alleen voor in HKS. Van deze personen heeft 95% tot 100% de codering ‘buitenland’ op de variabele GEM_CAT, hetgeen er op duidt dat het hier waarschijnlijk om staandehoudingen door de Koninklijke Marechaussee gaat.

Burgerlijke Staat: hier is gekozen voor (i) ongehuwd, onbekend (ii) gehuwd, partnerschap (iii) gescheiden, verweduwd.

Verblijfsduur: hier is gekozen voor (i) 0-5 (ii) 5+. Dit is verschillend van de codering in de vorige rapport, waar de categorieën (i) 0-5 (ii) 5-18 en (iii) 18+ zijn gehanteerd. Met het huidige GBA-bestand was dit niet mogelijk, aangezien de gegevens in de bestanden die het CBS voor statistische analyse prepareert maximaal teruggaan tot 01-01-1995.

Som antecedenten: hier is gekozen voor (i) 1 (ii) 2+.

Harddruggebruiker: we merken hier op dat de validiteit van deze variabele niet optimaal is. Het is aannemelijk dat soms nagelaten is dit kenmerk te registreren.

Appendix 2: mogelijke koppelingsfouten

Voor de huidige data gelden dezelfde risico's op koppelingsfouten als voor de data uit het voorgaande rapport (Van der Heijden, Zwane en Hessen, 2006). Daarom is de beschrijving uit het voorgaand rapport integraal overgenomen.

Er zijn twee soorten koppelingsfouten:

(i) koppelingen die hadden moeten plaatsvinden maar niet hebben plaatsgevonden

(ii) koppelingen die niet hadden moeten plaatsvinden maar hebben plaatsgevonden.

Voor de GBA als geheel – dus niet alleen op dat deel dat betrekking heeft op Antillianen - is naar fouten van het type (i) onderzoek verricht.

In de GBA hebben mensen een persoonsgebonden administratief nummer (A-nummer)⁶. Bij aanhouding door de politie wordt aan de hand van bijvoorbeeld de gegevens naam, adres en woonplaats het A-nummer opgezocht en meestal gevonden. Voor de periode 2000-2002 kon 89,6 % van alle HKS records gekoppeld worden op A-nummer en het aantal fouten dat hierbij optreedt is verwaarloosbaar.

Vervolgens is 3,1 % gekoppeld op geboortedatum, geslacht en adres. Van deze 3,1 % wordt ten hoogste 1 % fout gekoppeld (Arts et al, 2000), dus dit leidt tot een vertekening van maximaal 0,031 %.

Voor wat betreft gemiste koppelingen:

1,1 % van de HKS records heeft een adres in het buitenland (toeristen of mensen op familiebezoek) en hier kan dus geen sprake zijn van een gemiste koppeling.

Bij 0,6 % is het woonland onbekend, hier kan mogelijk sprake zijn van een gemiste koppeling maar deze kans lijkt gering;

Bij 0,2 % is de gemeente onbekend, hier kan sprake zijn van een gemiste koppeling;

Bij 5,5 % is om een andere reden niet gekoppeld. Deze groep bevat mensen waar dit onderzoek om is begonnen (langdurig verblijvend in Nederland niet geregistreerd in de GBA), maar mogelijk ook toeristen of mensen die hier op familiebezoek komen en een adres in Nederland opgeven. In deze laatste gevallen is niet zozeer sprake van een gemiste koppeling maar behoren de personen niet tot de doelgroep van dit onderzoek (langdurig verblijvend in Nederland maar niet geregistreerd in GBA). Tenslotte kan administratieve vervuiling deel uit maken van deze 5,5 %.

Voor de gekoppelde GBA-gegevens is als peildatum de laatste vrijdag van september gekozen, terwijl voor het HKS het gehele jaar geldt. Dit kan voor elk van de jaren tot gevolg gehad hebben dat er personen zijn die alleen in het HKS voorkomen maar niet in de GBA, omdat zij zich op een later tijdstip dan de peildatum in de GBA hebben ingeschreven. Hierdoor kunnen de schattingen enigszins te hoog uitvallen. Voor 2003 geldt dit validiteitsprobleem in sterkere mate omdat geen rekening gehouden kon worden met personen die zich later dan 31 december 2003 hebben ingeschreven in de GBA.

Wij lichten dit nader toe. De koppeling tussen HKS en GBA en toevoeging van het RIN-nummer (dat koppelingskenmerk met GBA-gegevens is) is in één keer uitgevoerd voor een cumulatief GBA-bestand 1995-2003 en een cumulatief HKS-bestand 1999-2004. Dat betekent dat aan iemand die bijv. alleen in 1999 in HKS voorkwam en zich pas in 2003 in GBA heeft ingeschreven, toch een RIN-nummer is toegekend. In dit rapport is de selectie of iemand wel of niet in beide bestanden zit, gemaakt op basis van het RIN nummer. Dit heeft de volgende consequenties, die wij aan de hand van een aantal vragen en antwoorden bespreken:

- a) *Stel iemand is in 2002 in HKS bekend en in november 2002 in GBA. Zit deze persoon dan, in 2002, zowel HKS als GBA?* Aannemende dat die persoon vanaf november 2002 in de GBA zit (en niet daarvoor), dan hoort hij niet tot de gekoppelde van GBA en HKS naar peildatum. Omdat de selectie van wel of niet in beide zit is gemaakt op basis van RIN-nummer zit de persoon vrijwel

⁶ Invoering van het BSN als persoonsgebonden nummer vond plaats in 2007. Daarvoor konden afnemers van gegevens uit de GBA indien zij een persoonsgebonden nummer nodig hadden, met name gebruik maken van het zogeheten A-nummer uit de GBA, het administratieve nummer.

zeker bij de mensen die in 2002 zowel in GBA als in HKS zitten, maar zijn de exclusieve GBA-variabelen (zoals burgerlijke staat) voor die persoon niet bekend.

- b) *Als a: is deze persoon in 2003 zowel in HKS als GBA? of alleen in GBA?* De persoon in kwestie zit wel in 2003 in de GBA (ook in het bestand naar peildatum tenzij hij zich weer uitschrijft of overlijdt). Die persoon komt alleen dan ook in HKS 2003 voor wanneer hij opnieuw verdachte is van een misdrijf. Als hij in 2003 niet opnieuw wordt geregistreerd, zit hij dus in 2003 alleen in de GBA.
- c) *Stel iemand zit in 2002 in HKS en in februari 2003 in GBA. Deze persoon zit dan in 2002 alleen in HKS. In 2003 in HKS en GBA, of alleen in GBA?* Aannemende dat die persoon vanaf februari 2003 in de GBA zit (en niet daarvoor), geldt voor 2003 hetzelfde als bij b): alleen wanneer die persoon weer verdachte wordt zal hij opduiken in de HKS. Omdat de selectie van wel of niet in beide zitten is gemaakt op basis van RIN-nummer, zit de persoon vrijwel zeker bij de mensen die in 2002 zowel in GBA als in HKS zitten, maar zijn de exclusieve GBA-variabelen (zoals burgerlijke staat) voor die persoon niet bekend.
- d) *Stel iemand zit in 2003 in HKS en in november 2003 in GBA. Deze persoon is dan in de data bekend als, in 2003, alleen HKS?* Aannemende dat die persoon vanaf november 2003 in de GBA zit (en niet daarvoor), dan hoort hij niet tot de gekoppelden van GBA en HKS beschreven naar peildatum. Omdat de selectie van wel of niet in beide zitten is gemaakt op basis van RIN-nummer zit de persoon vrijwel zeker wel bij de mensen die in 2003 zowel in GBA als in HKS zitten, maar zijn de exclusieve GBA-variabelen (zoals burgerlijke staat) voor die persoon niet bekend.

Tenslotte moet worden opgemerkt dat, mede naar aanleiding van dit onderzoek, het inzicht in de validiteit van de gekoppelde data zelf nog toeneemt. Mijn indruk is dat zicht op de uiteindelijke validiteit nog niet is uitgekristalliseerd

Appendix 3: 'Capture-recapture' analyse⁷

De naam 'capture-recapture' (vangst-hervangst) komt oorspronkelijk uit de biologie waar deze methode wordt gebruikt bij het schatten van populatiegroottes van dieren. Hierbij wordt gepoogd dergelijke dieren herhaaldelijk te vangen. Een van de eerste voorbeelden van de 'capture-recapture' methode behelst het schatten van de hoeveelheid Noord-Amerikaanse eenden (Seber, 1982). Lincoln ving en ringde grote aantallen eenden voordat deze uitgezet werden op broedplaatsen.

Om de omvang N van een bepaalde populatie dieren te schatten neemt men een willekeurige steekproef van n_1 dieren uit deze populatie van N dieren. Deze n_1 dieren worden van een merkteken voorzien en weer uitgezet. Na verloop van tijd (zodat de gemerkte dieren zich goed hebben kunnen mengen met de niet-gemerkte dieren) neemt men een tweede willekeurige steekproef van n_2 dieren. Hiervan blijken er m_2 een merkteken te hebben. Op basis hiervan kan N geschat worden: als we aannemen dat de proportie gemerkten in de tweede steekproef (d.w.z. m_2 / n_2) een perfect beeld geeft van de proportie gemerkten in de populatie (d.w.z. n_1 / N), dan geldt dat $m_2 / n_2 = n_1 / N$, en dus schatten we N met $n_1 n_2 / m_2$. Hierbij is gebruik gemaakt van de veronderstelling dat de twee steekproeven statistisch onafhankelijk van elkaar zijn.

Bovenstaande aanpak is typisch voor de biologie, waar men meer dan één steekproef trekt uit dezelfde populatie. In de epidemiologie en in het gebied dat men wel 'official statistics' noemt, trekt men zelden specifiek voor dit doeleinde meerdere steekproeven, maar maakt men eerder gebruik van bestaande registraties en onderzoekt de overlap tussen deze registraties. Dat is ook gebeurd in de schatting van het aantal Antillianen dat niet in de GBA staat geregistreerd. We zullen aan de hand een epidemiologisch onderzoek, namelijk het tweede incidentie onderzoek van diabetes mellitus bij 0-19 jarigen, laten zien wat de oplossing is die de 'capture-recapture' methode levert.

Voor de berekening van de incidentie van diabetes beschikte men over twee registraties, namelijk die uit een enquête onder alle kinderartsen en internisten in Nederland (verder afgekort KIN) en die van de Diabetes Vereniging Nederland (DVN) (Hiraseng, 1995). Wij concentreren ons hier op de schatting van de incidentie bij 0-4 jarigen van 1988-1990, en refereren naar deze groep als een populatie. Er waren 160 kinderen bekend bij KIN, en 110 bij de DVN. De overlap tussen beide registraties was 99. Gebruik van de registratie van de DVN leverde dus 11 nieuwe kinderen op. We kunnen de volgende tabel maken.

	niet in DVN	in DVN	Totaal
niet in KIN	?	11	11 + ?
in KIN	61	99	160
Totaal	61 + ?	110	171 + ?

Als we KIN beschouwen als de eerste steekproef en DVN als de tweede, is $n_1 = 160$, $n_2 = 110$, $m_2 = 99$, en $N = 171 + ?$. De hierboven beschreven methode levert als schatting van de populatieomvang N van 0-4 jarigen met diabetes mellitus $n_1 n_2 / m_2 = 177,8$ op, en het vraagteken (het aantal kinderen dat in géén van beide registraties zit) is dus 6,8 (95 % betrouwbaarheidsinterval 3,5 – 13,2).

Bij de schatting van de Antillianen is de interesse niet zozeer uitgegaan naar de schatting van de gehele populatie Antillianen, maar slechts naar de Antillianen die niet geregistreerd stonden in de GBA. Indien we in de tabel hierboven KIN vervangen door GBA, en DVN door HKS, dan gaat de interesse uit naar het aantal 11 en het vraagteken, dat dus geschat is is 6,8.

Voor een discussie van de vooronderstellingen van de vangst-hervangst methode verwijzen we naar IWGDMF (1995) en Van der Heijden en Hiraseng (2002). Voor de vooronderstellingen bij de schatting van het aantal Antillianen dat niet in de GBA is geregistreerd verwijzen we naar de hoofdstekst.

⁷ Deze appendix is een ingekorte en gewijzigde versie van Van der Heijden en Hiraseng (2002). Een uitgebreid overzicht van dit type methoden is ook te vinden in Sikkels, Van der Heijden en Van Gils (2006).

In potentie kan een variabele die invloed heeft op de kans van voorkomen in zowel de GBA als in HKS een enorme invloed hebben. Hier staat tegenover dat als een variabele invloed heeft op slechts één van beide kansen, dit niet bezwaarlijk is. Om deze reden hebben wij zoveel mogelijk gebruik gemaakt van extra variabelen, zie Appendix A. De wijze waarop wij dit hebben gedaan, staat beschreven in Zwane en Van der Heijden (2007) en Van der Heijden, Zwane en Hessen (2009). Hier staat ook beschreven hoe de betrouwbaarheidsintervallen van de schattingen worden geschat.

Referenties

- Arts, Koos, Bart F.M. Bakker, en Erik van Lith (2000). Linking administrative registers and household surveys, In: Pieter Al en Bart F.M. Bakker (eds.), *Re-engineering Social Statistics by micro-integration of different sources*. Themanummer Netherlands Official Statistics, jrg. 15, nr. summer, blz. 16-22
- CBS (2005). *Rapportage koppeling HKS2004 – CKP*. Voorburg: CBS
- Chao, A., Tsay, P., Lin, S., Shau, W., and Chao, D. (2001) The applications of capture-recapture models to epidemiological data. *Statistics in Medicine*, 20, 3123--3157.
- Van der Heijden, P.G.M. en R. Hirasing (2002). De vangst-hervangstmethode. In: L.M. Bouter en F.R. Rosendaal. *Dwalen in de methodologie. Een bundel artikelen uit het Nederlands Tijdschrift voor de Geneeskunde*. Amsterdam: Nederlands Tijdschrift voor geneeskunde. Pp. 84-87.
- Van der Heijden, P.G.M., Zwane, E., & Hessen, D. (2006). Schatting van aantal in Nederland verblijvende Antillianen die niet ingeschreven zijn in de GBA. Een 'capture-recapture'-analyse in opdracht van het Ministerie van Justitie. Utrecht, Universiteit Utrecht, IOPS. (22 pagina's).
- Van der Heijden, P.G.M., E. Zwane & D. Hessen (2009). Structurally missing data problems in multiple list capture-recapture data. *Advances of Statistical Analysis*, 93, 5-21.
- Hirasing, R.A. (1995). De incidentie van diabetes mellitus bij 0-19 jarigen in Nederland (1988-1990). *Ned Tijdschr Geneesk*, 139, 1088-91.
- International Working Group for Disease Monitoring and Forecasting (1995). Capture-recapture and multiple record systems estimation 1: history and theoretical development. *Am J Epidemio*. 142, 1047-58.
- Seber, G.A.F. (1982). *The estimation of animal abundance and related parameters*. London: Griffin.
- Sikkel, D., Van der Heijden, P.G.M., & Van Gils, G. (2006). *Methoden voor omvangschattingen van verborgen populaties, met name illegalen*. (WODC rapport 243, Onderzoek en beleid). Meppel: Boom Juridische Uitgevers.
- Van der Vliet, R. (ongedateerd). *Beschrijvingen variabelen van de Integratiekaart-bestanden*. Voorburg: CBS.
- Zwane, E., & Van der Heijden, P.G.M. (2007). Analysing capture-recapture data when some variables of heterogeneous catchability are not collected or asked in all registrations. *Statistics in medicine*, 26, 1069-1089

De publicaties met Van der Heijden als (co)auteur zijn te downloaden van zijn homepage www.fss.uu.nl/ms/vanderheijden .

Bijlage 1. Aantallen en percentages Antillianen uitgesplitst naar geslacht en leeftijd. Achtereenvolgens: Geschat aantal niet in GBA; Aantal in GBA; Totaal aantal (som eerdere twee kolommen); Percentage niet in GBA (= Niet in GBA/ Totaal); Percentage bekend in HKS (percentage bepaald voor zowel bekend als niet bekend in GBA); Marge: met bootstrap berekend 95 % betrouwbaarheidsinterval voor Niet in GBA (percentielen 2,5 %, 50 % en 97,5 %, merk op dat betrouwbaarheidsinterval niet symmetrisch is).

2004

		Niet in GBA	in GBA	Totaal	% niet in GBA	% in HKS	Marge niet in GBA		
							2,5 %	50 %	97,5 %
<i>Man</i>	<i>13-17</i>	66.0	2529	2595.0	2.5	18.6	33.4	64.8	105.6
<i>Man</i>	<i>18-24</i>	245.0	6455	6700.0	3.7	19.4	176.9	243.6	324.4
<i>Man</i>	<i>25-44</i>	382.6	17913	18295.6	2.1	17.3	295.4	380.0	478.0
<i>Man</i>	<i>45+</i>	271.0	10933	11204.0	2.4	7.6	147.7	269.9	391.7
<i>Vrouw</i>	<i>13-17</i>	67.0	2324	2391.0	2.8	4.6	0.0	65.0	154.4
<i>Vrouw</i>	<i>18-24</i>	291.0	6693	6984.0	4.2	5.3	155.8	293.6	465.8
<i>Vrouw</i>	<i>25-44</i>	508.6	17362	17870.6	2.8	4.4	308.3	508.5	743.3
<i>Vrouw</i>	<i>45+</i>	550.1	13203	13753.1	4.0	1.3	205.6	544.0	1051.8
Totaal		2381.3	77412	79793.3	3.0	9.1	1919.2	2373.6	3007.2

2005

		Niet in GBA	in GBA	Totaal	% niet in GBA	% in HKS	Marge niet in GBA		
							2,5 %	50 %	97,5 %
<i>Man</i>	<i>13-17</i>	24.8	2302	2326.8	1.1	20.4	5.0	24.1	50.2
<i>Man</i>	<i>18-24</i>	175.2	6143	6318.2	2.8	19.3	122.4	174.4	242.0
<i>Man</i>	<i>25-44</i>	316.7	17376	17692.7	1.8	17.0	238.3	314.8	403.6
<i>Man</i>	<i>45+</i>	252.8	11388	11640.8	2.2	7.7	147.5	254.1	368.4
<i>Vrouw</i>	<i>13-17</i>	0.0	2168	2168.0	0.0	7.2	0.0	0.0	0.0
<i>Vrouw</i>	<i>18-24</i>	108.0	6158	6266.0	1.7	4.7	21.8	108.2	219.7
<i>Vrouw</i>	<i>25-44</i>	268.5	16669	16937.5	1.6	4.2	120.4	268.1	434.8
<i>Vrouw</i>	<i>45+</i>	387.8	13650	14037.8	2.8	1.3	77.9	369.9	771.4
Totaal		1533.7	75854	77387.7	2.0	9.0	1132.5	1529.8	1993.3

- Bijlage 1 vervolg-

2006

		Niet in GBA	in GBA	Totaal	% niet in GBA	% in HKS	Marge 2,5 %	niet in GBA 50 %	97,5 %
<i>Man</i>	<i>13-17</i>	5.8	2199	2204.8	0.3	17.4	0.0	5.7	18.1
<i>Man</i>	<i>18-24</i>	82.2	5883	5965.2	1.4	18.4	43.6	80.6	125.4
<i>Man</i>	<i>25-44</i>	187.5	17039	17226.5	1.1	16.1	125.2	186.2	260.0
<i>Man</i>	<i>45+</i>	145.4	11883	12028.4	1.2	7.7	66.4	143.3	240.7
<i>Vrouw</i>	<i>13-17</i>	0.0	2080	2080.0	0.0	6.1	0.0	0.0	0.0
<i>Vrouw</i>	<i>18-24</i>	40.5	5910	5950.5	0.7	5.0	0.0	39.1	104.4
<i>Vrouw</i>	<i>25-44</i>	91.3	16257	16348.3	0.6	4.4	21.8	89.8	199.6
<i>Vrouw</i>	<i>45+</i>	287.7	14169	14456.7	2.0	1.4	64.3	279.1	631.8
Totaal		840.4	75420	76260.4	1.1	8.7	539.8	828.6	1208.1

2007

		Niet in GBA	in GBA	Totaal	% niet in GBA	% in HKS	Marge 2,5 %	niet in GBA 50 %	97,5 %
<i>Man</i>	<i>13-17</i>	11.3	2133	2144.3	0.5	17.8	0.0	11.0	29.0
<i>Man</i>	<i>18-24</i>	117.1	5830	5947.1	2.0	19.1	67.8	116.2	167.1
<i>Man</i>	<i>25-44</i>	234.3	16914	17148.3	1.4	16.0	163.7	236.3	318.6
<i>Man</i>	<i>45+</i>	26.8	12523	12549.8	0.2	7.5	0.0	26.4	68.0
<i>Vrouw</i>	<i>13-17</i>	35.3	2066	2101.3	1.7	5.8	0.0	33.3	91.7
<i>Vrouw</i>	<i>18-24</i>	93.7	5868	5961.7	1.6	5.4	19.0	92.7	190.7
<i>Vrouw</i>	<i>25-44</i>	174.1	15996	16170.1	1.1	4.1	51.2	169.7	320.8
<i>Vrouw</i>	<i>45+</i>	328.3	14772	15100.3	2.2	1.2	64.4	322.1	696.2
Totaal		1021.0	76102	77123.0	1.3	8.5	667.2	1007.0	1400.6