

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**

Peter van der Heijden, Eugene Zwane en Dave Hessen
Februari 2006
IOPS-Utrecht
Universiteit Utrecht

Wij bedanken de heren dr. B.F.M. Bakker en drs. J. Oudhof (beiden CBS) voor het uitvoeren van de koppeling van de GBA en het HKS, het leveren van de bestanden waarop wij onze berekeningen hebben uitgevoerd, de adviezen die zij tijdens het onderzoek hebben gegeven en het becommentariëren van dit manuscript. Zij zijn echter op generlei wijze verantwoordelijk voor de inhoud van dit manuscript.

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

Peter G.M. van der Heijden and Eugene Zwane

Samenvatting

In dit rapport presenteren wij schattingen van het aantal Antillianen dat in Nederland verblijft zonder ingeschreven te zijn in de Gemeentelijke Basis Administratie (GBA). Deze schattingen lopen van ongeveer 3.000 in het jaar 2000 tot minder dan 12.000 in het jaar 2003.

De schattingen zijn 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 vier groepen Antillianen te onderscheiden: (i) de groep die zowel in GBA als in HKS bekend is; (ii) de groep die wel in HKS maar niet in GBA bekend is; (iii) de groep die wel in GBA maar niet in HKS bekend is; (iv) de groep die noch in GBA noch in HKS bekend is. Als een tussenresultaat levert de capture-recapture methode een schatting van de omvang van de groep genoemd onder (iv). Door de omvang van groepen (ii) en (iv) op te tellen ontstaan bovengenoemde schattingen. Er zijn uitsplitsingen van de schattingen naar leeftijd, geslacht, het al dan niet hebben van antecedenten, het al dan niet zijn van harddruggebruiker, verblijfsduur en burgerlijke staat. Er worden ook schattingen gepresenteerd 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.

Estimation of the size of the population of Dutch Antilleans staying in the Netherlands without being officially registered; A capture-recapture analysis for the Ministry of Justice

Peter G.M. van der Heijden and Eugene Zwane

Summary

In this report we present estimates of the size of the population of Dutch Antilleans that stay in the Netherlands without being officially registered. These estimates range from around 3,000 in the year 2000 to less than 12,000 in the year 2003.

These estimates are based on an application of the so-called capture-recapture method. The registrations are linked by Statistics Netherlands, namely an automated system of decentralised (municipal) population registers (Dutch abbreviation: GBA) and a police registration (Dutch abbreviation: HKS). By linking these registrations we can define four groups of Dutch Antilleans: (i) the group known in GBA as well as in HKS (ii) the group known in HKS but not in GBA, (iii) the group known in GBA but not in HKS, and (iv) the group known in neither of the HKS or the GBA. As an intermediate result the capture-recapture method provides an estimate of the size of the group defined in (iv). By adding up the sizes of the groups under (ii) and (iv) we find the estimates provided above. The estimates are split up over levels of the variables gender, age, formerly being known by the police, being a hard drug user, length of stay in the Netherlands and marital status. We also provide estimates of percentages of each group known in HKS.

The capture-recapture method uses a number of assumptions at arrive at estimates. The tenability of these assumptions is discussed in detail

1. Inleiding

In de Notitie Antilliaanse Risicjongeren (Minister van V&I, dd 4 oktober 2004) wordt aangegeven dat er signalen zijn dat er Antilliaanse jongeren zijn die niet of niet correct in de Gemeentelijke Basis Administratie (GBA) geregistreerd staan, waarschijnlijk doordat zij zich, eenmaal in Nederland gevestigd, niet inschrijven in de GBA van hun woonplaats of, als zij zich wel hebben ingeschreven, zij feitelijk niet woonachtig zijn op het adres van inschrijving. Bekend is dat ze regelmatig pendelen tussen Nederland en de Nederlandse Antillen en ook binnen Nederland zeer mobiel zijn.

Voor een effectieve aanpak van o.a. criminaliteit is het van belang dat consensus bestaat over de oorzaken waarom jongeren moeilijk binnen het bereik van gemeenten komen. Om op dit punt over meer exacte gegevens te beschikken heeft Justitie eerst – via het onderhavige onderzoek – door ons een quick scan laten uitvoeren naar de grootte van de groep die niet staat ingeschreven in de bevolkingsadministratie. De ‘waarom-vraag’ zal in aanvullend onderzoek moeten worden onderzocht.

Wij zijn benaderd door het WODC met het verzoek om een schatting te maken van het aantal ‘illegaal’ in Nederland verblijvende Antillianen. Doel is de schatting van het aantal niet-geregistreerde Antillianen in Nederland. Wij merken hierbij op dat het gaat om zowel criminele als niet-criminele Antillianen, en om zowel jonge als oude Antillianen.

Het schatten van het aantal niet-geregistreerde Antillianen zal op een objectieve wijze moeten plaatsvinden (dus niet via inschattingen van deskundigen). Door diverse registratiesystemen te koppelen is dit mogelijk. Kernvraag is: Wat is het aantal Antillianen (jongeren en ouderen) waarvoor geldt dat zij langere tijd in Nederland verblijven, zonder geregistreerd te zijn in de GBA?

2. Methode

De schatting zal plaatsvinden met moderne varianten van de zgn. capture-recapture methode. In deze schattingsmethode wordt gebruik gemaakt van gekoppelde gegevens. Hiertoe worden twee bestanden gekoppeld:

- GBA gegevens (verkregen met de Integratiekaart van het CBS)
- HKS gegevens.

Uit deze beide bestanden worden alle personen geselecteerd die als geboorteplaats de Nederlandse Antillen hebben opgegeven.

Na de koppeling is de volgende tabel te maken:

	Niet in GBA	In GBA
Niet in HKS	(1)	(2)
In HKS	(3)	(4)

In (4) zit het aantal mensen dat zowel in GBA als in HKS bekend is;

In (3) zit het aantal mensen dat wel in HKS maar niet in GBA bekend is;

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

In (1) zit het aantal mensen dat noch in GBA noch in HKS bekend is.

Het aantal mensen in (1) wordt door ons geschat. De optelling van deze schatting bij het aantal in (3) levert dan het antwoord op de onderzoeksvraag: hoeveel Antillianen zijn er die in Nederland aanwezig zijn 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 GBA zijn dit:

- Leeftijd
- Geslacht

Voor alleen HKS zijn dit:

- Som antecedenten
- Harddruggebruiker (ja/nee)

Voor alleen GBA zijn dit:

- Verblijfsduur
- Burgerlijke staat.

Voor meer informatie over deze variabelen verwijzen wij naar Appendix 1.

De door ons gebruikte statistische methoden staan beschreven in Appendix 3, Chao et al. (2001), en Zwane en van der Heijden (2004, 2005).

Nadere precisering van de onderzoeksvraag

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

- Het is gezien de wijze van registreren in de HKS en de GBA niet goed mogelijk om een onderscheid aan te brengen tussen Antillianen en Arubanen.
- Het is door de aard van de bestanden de daarin aanwezige gegevens niet mogelijk om bij dit onderzoek een onderscheid aan te brengen tussen verschillende generaties

waaruit Antillianen afkomstig zijn. De uitkomsten in dit onderzoek hebben slechts betrekking op eerste generatie Antillianen.¹

- Het is mogelijk dat een Antiliaan zich inschrijft bij een gemeente en hierbij een briefadres opgeeft als adres.² 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 aannamen gemaakt, die bij het maken van dergelijke schattingen gebruikelijk zijn. 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 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 (2004 of 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.

¹ Inhoudelijk zou het ook niet zo voor de hand liggen om onderzoek te doen naar de aanwezigheid van 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).

² Dit moet overigens wel een bestaand woonadres zijn (en kan dus niet een postbus 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 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.
- Indien deze aanname geschonden is doordat er een *negatieve* samenhang tussen deze kansen is, dat wil zeggen dat personen die zich *vaker* inschrijven bij 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 GBA) 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 aflight van de door ons gepresenteerde schatting. Wij weten echter wel de bovengrens, dus al 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 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. Het is niet duidelijk hoe ernstig een schending van deze veronderstelling is, en in welke richting de door ons geleverde schatting zal afwijken van de werkelijke populatieomvang. Wij zijn geneigd de invloed van deze mogelijke schending als gering in te schatten.

Dit brengt ons tot de conclusie dat, indien aannamen 1 en 3 zijn geschonden, dit heeft geleid tot te hoge schattingen. Onduidelijk is echter of er sprake is van schendingen van de aannamen.

Wij willen tot slot opmerken dat bij elke statistische analyse aannamen worden gemaakt, maar deze meestal niet worden geëxpliciteerd, zoals wij hier wel doen. Daarnaast betekent een schending van een aanname niet noodzakelijk dat er een grote invloed op de uitkomsten hoeft te zijn. Schattingen kunnen soms vrij robuust zijn voor schending van aannamen. De robuustheid van de schattingen is te onderzoeken door gevoeligheidsanalyses te doen, maar dit past niet binnen de 'quick scan' die het Ministerie voor ogen staat. Als laatste merken we op dat de door ons gebruikte methoden standaard zijn voor het doen van populatieschattingen. Maar dit betekent niet dat deze schattingen er nooit 'naast' zullen zitten.

3. Resultaten

3.1 Enkele bevindingen vooraf

Hieronder vindt u de marginale frequenties op basis van de GBA en HKS. Enkele zaken vallen hier op:

- er is een groei van het aantal in GBA ingeschrevenen van ongeveer 69.000 in 2000 tot 79.000 in 2003.
- Er is een veel grotere groei van het aantal personen dat *niet* in de GBA bekend is maar wel in aanraking is gekomen met de politie, van 201 in 2000 tot 949 in 2003.
- De te schatten aantallen zullen van 2000 tot 2003 om de volgende reden fors stijgen: de aanname van onafhankelijkheid tussen opname in GBA en opname in HKS betekent dat, bijvoorbeeld voor het jaar 2000, $?/201 = 64.247/5.095$, zodat $? = 201 * 64.247/5.095$. Omdat dit aantal van 201 stijgt tot 949, stijgt de schatting van het vraagteken ook met ongeveer een factor 5.

Wij komen hier nog niet met gedetailleerde schattingen, omdat wij schattingen willen controleren voor variabelen als leeftijd en geslacht.

Wij merken hier overigens op dat de steekproeven voor achtereenvolgende jaren grotendeels gebaseerd zijn op identieke informatie, omdat, bijvoorbeeld, als iemand in 2001 in GBA zit, deze persoon ook meestal in 2000, 2002 en 2003 in GBA zit. Er is dus niet sprake van onafhankelijke steekproeven, en de toetsresultaten over de jaren heen zijn niet onafhankelijk van elkaar. Wij hebben daarom niet een simultane analyse uitgevoerd op alle vier jaren tegelijkertijd.

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

3.2 Schattingen gecontroleerd voor leeftijd en geslacht

In een eerste reeks analyses hebben wij onderzocht of Leeftijd, Geslacht en Regio van invloed zijn op de kans niet in GBA te worden aangetroffen. De effecten die we voor Regio aantreffen waren significant maar zeer gering. Daarom hebben we besloten om deze variabele verder niet in de analyses mee te nemen.

De eerste analyseresultaten zijn die voor het model waarin de kans om in 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.

De onderstaande tabellen geven de belangrijkste informatie weer. We leggen hier de eerste regel uit:

- niet in GBA wil zeggen: het aantal personen dat in HKS zit maar niet in GBA (dus in cel (3)) plus het geschatte aantal personen dat noch in HKS noch in GBA zit (cel (1)). Voor jongens van 13-17 zijn dit er in 2000 dus 41.5.
- in GBA wil zeggen: het aantal personen dat in GBA zit, al dan niet in HKS (dus cel (2) en (4) samen). Voor jongens van 13-17 zijn dit er in 2000 dus 3.168.5.
- Totaal wil zeggen: de som van de eerste twee getallen, dus het totale aantal Antillianen dat in Nederland verblijft. Voor jongens van 13-17 zijn dit er in 2000 dus 3.185,2.
- % niet in GBA wil zeggen: het eerste getal gedeeld door het totaal. Voor jongens van 13-17 zijn dit er in 2000 dus $100 * 41.5 / 3168,5 = 1.3 \%$.
- % in HKS: is berekend door de getallen (1) en (2) te delen door de som van de getallen (1), (2), (3) en (4). Voor jongens van 13-17 komt in 2000 dus 12,1 % voor in HKS.
- Marge niet in GBA: geeft een 95 % betrouwbaarheidsinterval aan voor het eerste getal, bepaald met de bootstrap. Het aantal van 41.5 heeft dus een 95 % betrouwbaarheidsinterval dat loopt van 8,1 tot 81,7. Met deze gegevens zijn, desgewenst, ook de betrouwbaarheidsgrenzen te bepalen van de 1.3 %, namelijk als $8,1 / (8,1 + 3127)$ en $81,7 / (81,7 + 3127)$.

Men kan allerlei informatie uit de tabellen halen. Wij noemen wat zaken die ons als beleidsmatig interessant voorkomen:

- wij schatten het aantal niet geregistreerde personen op ongeveer 3.000 in het jaar 2000, 6.300 in 2001, 8.900 in 2002 en 11.800 in 2003. Het niet registreren in GBA komt steeds vaker voor: van 2000 tot 2003 is er een stijging ten opzichte van 4 % tot 13 %.
- in alle leeftijdscategorieën komt het niet registreren voor, en zowel bij mannen als vrouwen. De leeftijdscategorie waarin niet registreren het vaakst voorkomt is 18-24.
- Wij merken op dat het hier schattingen betreft die marges hebben. De marge van de schattingen is aanzienlijk: het aantal in GBA gemiste personen in 2000 is (afgerond) 3000, en dit ligt tussen 2.400 en 3.700 (d.w.z. 20% boven en onder). Relatief gezien nemen de marges in latere jaren af (tot ongeveer 10 % boven en onder 11.700 in 2003). We merken ook op dat de marges voor de groepen 45+ groter zijn, omdat daar de ophoogfactor groter is.
- Hoewel dit niet tot de onderzoeksvraag behoort, hebben wij ook percentages gegeven van de Antillianen in Nederland die in HKS voorkomen. Deze zijn laag bij de vrouwen (onder de 5 %) en hoger bij de mannen, vooral in de leeftijd 18-24. Wij merken op dat deze percentages identiek zijn onder de Antillianen die geregistreerd zijn in GBA en zij die dit niet zijn, omdat dit een van de aannamen van onze analyses is.

3.3 Schattingen gecontroleerd voor leeftijd, geslacht, burgerlijke staat, verblijfsduur, aantal antecedenten en het al dan niet zijn van druggebruiker

Met statistische methoden beschreven in Zwane en van der Heijden (2005) hebben wij schattingen berekend waarbij gecontroleerd wordt voor de variabelen leeftijd, geslacht, burgerlijke staat, verblijfsduur, aantal antecedenten en het al dan niet zijn van druggebruiker. Hierbij is het volgende loglineaire model geschat: $(gba + hks) * (geslacht * leeftijd) + hks * burgerlijke\ staat * verblijfsduur + gba * antecedenten * druggebruiker + (burgerlijke\ staat + verblijfsduur) * (antecedenten + druggebruiker)$. Wij merken hierbij op dat burgerlijke staat en verblijfsduur variabelen zijn die alleen in GBA voorkomen, en antecedenten en druggebruiker alleen in HKS.

Wij geven de schattingen weer voor de volgende combinaties van variabelen:

- leeftijd en geslacht
- burgerlijke staat en verblijfsduur
- antecedenten en druggebruiker.

Voor leeftijd en geslacht is het patroon in Tabel 2 vergelijkbaar aan het patroon in Tabel 1.

Voor burgerlijke staat zien we dat de percentages niet in GBA ingeschrevenen hoger zijn voor ongehuwden, vervolgens voor de gescheidenen, en de percentages zijn het laagst voor de gehuwden. Hoe korter de verblijfsduur, des te groter het percentage niet ingeschrevenen in GBA.³ We zien een identiek patroon voor het bekend zijn in HKS.

De personen met slechts een enkel antecedent komen vaker niet voor in de GBA dan personen met twee of meer antecedenten. Harddruggebruikers zijn vrijwel allemaal in de GBA bekend.

³ Inhoudelijk gezien komt deze uitspraak mogelijk vreemd over, omdat verblijfsduur alleen bekend is bij de personen die ingeschreven zijn in de GBA. Punt is echter dat wij in onze schattingsprocedures personen bijschatten die niet in de GBA voorkomen, en daarvoor tegelijkertijd de verblijfsduur bijschatten.

Tabel 1. Aantallen en percentages Antillianen uitgesplitst naar geslacht en leeftijd. Kolom 1: Geschat aantal niet in GBA. Kolom 2: aantal in GBA. Kolom 3: Totaal aantal Antillianen (som kolom 1 en 2). Kolom 4: percentage Antillianen niet in GBA (= kolom 1 / kolom 3). Kolom 5: percentage Antillianen bekend in HKS (percentage bepaald voor zowel Antillianen bekend als niet bekend in GBA), Kolom 6,7,8: betrouwbaarheidsinterval voor kolom 1.

2000

		Niet in GBA	in GBA	Totaal	% niet in GBA	% in HKS	Marge niet in GBA		
							2,5 %	50 %	97,5 %
<i>Man</i>	13-17	41,5	3127	3168,5	1,3	12,1	8,1	41,2	81,7
<i>Man</i>	18-24	272,0	6101	6373,0	4,3	16,2	193,2	270,4	355,1
<i>Man</i>	25-44	662,1	16826	17488,1	3,8	13,9	521,8	659,5	802,6
<i>Man</i>	45+	296,9	7775	8071,9	3,7	5,4	158,2	292,7	461,9
<i>Vrouw</i>	13-17	106,7	3058	3164,7	3,4	2,8	0,0	100,7	255,0
<i>Vrouw</i>	18-24	317,3	6210	6527,3	4,9	4,4	149,7	313,8	514,6
<i>Vrouw</i>	25-44	678,6	16441	17119,6	4,0	3,2	401,2	677,8	965,3
<i>Vrouw</i>	45+	605,2	9804	10409,2	5,8	0,8	127,2	574,9	1248,9
Totaal		2980,2	69342	72322,2	4,1	7,3	2393,9	2952,5	3703,8

2001

		Niet in GBA	in GBA	Totaal	% niet in GBA	% in HKS	Marge niet in GBA		
							2,5 %	50 %	97,5 %
<i>Man</i>	13-17	87,8	3294	3381,8	2,6	13,7	42,3	86,8	136,9
<i>Man</i>	18-24	652,5	6583	7235,5	9,0	15,6	530,5	657,9	783,0
<i>Man</i>	25-44	1237,2	17903	19140,2	6,5	13,7	1059,8	1236,1	1437,5
<i>Man</i>	45+	479,4	8725	9204,4	5,2	5,2	304,4	482,3	694,7
<i>Vrouw</i>	13-17	196,8	3236	3432,8	5,7	4,6	82,9	194,8	354,5
<i>Vrouw</i>	18-24	753,7	6822	7575,7	9,9	5,2	514,8	745,0	1006,9
<i>Vrouw</i>	25-44	1566,1	17524	19090,1	8,2	3,7	1154,7	1561,5	2032,5
<i>Vrouw</i>	45+	1336,8	10900	12236,8	10,9	1,0	692,6	1326,2	2212,7
Totaal		6310,3	74987	81297,3	7,8	7,5	5448,3	6333,7	7387,2

-tabel 1 vervolg-

2002

		Niet		Totaal	% niet		Marge niet in GBA		
		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
<i>Man</i>	<i>13-17</i>	158,5	3328	3486,5	4,5	15,1	99,3	156,9	226,0
<i>Man</i>	<i>18-24</i>	1165,9	6828	7993,9	14,6	17,9	1013,4	1165,5	1333,0
<i>Man</i>	<i>25-44</i>	2055,9	18441	20496,9	10,0	16,1	1842,0	2067,4	2296,3
<i>Man</i>	<i>45+</i>	784,8	9483	10267,8	7,6	6,1	579,3	798,1	1034,7
<i>Vrouw</i>	<i>13-17</i>	188,3	3222	3410,3	5,5	4,8	69,2	183,4	328,6
<i>Vrouw</i>	<i>18-24</i>	1007,5	7113	8120,5	12,4	5,0	736,6	1012,4	1304,4
<i>Vrouw</i>	<i>25-44</i>	2612,0	17929	20541,0	12,7	3,9	2116,7	2612,2	3216,1
<i>Vrouw</i>	<i>45+</i>	927,4	11771	12698,4	7,3	1,4	457,0	931,5	1487,6
Totaal		8900,3	78115	87015,3	10,2	8,5	8061,7	8932,2	9916,2

2003

		Niet		Totaal	% niet		Marge niet in GBA		
		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
<i>Man</i>	<i>13-17</i>	126,4	3180	3306,4	3,8	15,0	71,9	125,4	185,3
<i>Man</i>	<i>18-24</i>	1542,0	6788	8330,0	18,5	17,9	1349,4	1543,3	1750,0
<i>Man</i>	<i>25-44</i>	2042,1	18414	20456,1	10,0	16,9	1828,5	2043,3	2279,2
<i>Man</i>	<i>45+</i>	786,4	10182	10968,4	7,2	7,2	580,2	791,1	1010,7
<i>Vrouw</i>	<i>13-17</i>	161,9	3007	3168,9	5,1	4,3	60,7	159,5	300,5
<i>Vrouw</i>	<i>18-24</i>	1921,5	7102	9023,5	21,3	5,3	1532,3	1910,6	2357,8
<i>Vrouw</i>	<i>25-44</i>	2704,6	17875	20579,6	13,1	4,1	2167,7	2678,5	3255,6
<i>Vrouw</i>	<i>45+</i>	2498,4	12492	14990,4	16,7	1,2	1605,0	2495,4	3513,0
Totaal		11783,3	79040	90823,3	13,0	8,7	10576,9	11738,9	13132,3

Tabel 2. Aantallen en percentages Antillianen uitgesplitst naar geslacht en leeftijd. Kolom 1: Geschat aantal niet in GBA. Kolom 2: aantal in GBA. Kolom 3: Totaal aantal Antillianen (som kolom 1 en 2). Kolom 4: percentage Antillianen niet in GBA (= kolom 1 / kolom 3). Kolom 5: percentage Antillianen bekend in HKS (percentage bepaald voor zowel Antillianen bekend als niet bekend in GBA). Kolom 6,7,8: betrouwbaarheidsinterval voor kolom 1.

2000

Leeftijd		Niet in GBA	in GBA	Totaal	% niet in GBA	% in HKS	Marge niet in GBA		
							2,5 %	50 %	97,5 %
<i>Man</i>	13-17	39,4	3127	3166,4	1,2	12,1	7,9	37,9	77,8
<i>Man</i>	18-24	258,7	6101	6359,7	4,1	16,2	185,7	261,1	338,4
<i>Man</i>	25-44	629,0	16826	17455,0	3,6	13,9	508,7	629,0	766,1
<i>Man</i>	45+	280,6	7775	8055,6	3,5	5,4	152,5	274,3	421,6
<i>Vrouw</i>	13-17	100,7	3058	3158,7	3,2	2,8	0,0	100,4	245,2
<i>Vrouw</i>	18-24	299,7	6210	6509,7	4,6	4,4	152,9	298,4	460,0
<i>Vrouw</i>	25-44	640,5	16441	17081,5	3,7	3,2	400,2	654,1	934,6
<i>Vrouw</i>	45+	570,3	9804	10374,3	5,5	0,8	118,7	556,8	1175,3
Totaal		2818,9	69342	72160,9	3,9	7,3	2202,4	2827,4	3580,0

Verblijfsd.		Niet in GBA	in GBA	Totaal	% niet in GBA	% in HKS	Marge niet in GBA		
							2,5 %	50 %	97,5 %
<i>Ongehuwd</i>	0- 5	926,8	17230	18157,2	5,1	10,0	720,8	927,7	1172,7
<i>Ongehuwd</i>	5-18	763,4	19561	20324,2	3,8	9,0	597,2	765,9	969,1
<i>Ongehuwd</i>	18+	225,1	6750	6975,1	3,2	9,0	175,6	225,7	286,6
<i>Gehuwd</i>	0-5	154,2	3490	3644,6	4,2	3,7	119,5	154,2	197,6
<i>Gehuwd</i>	5-18	207,7	6237	6444,6	3,2	2,3	158,1	208,8	267,9
<i>Gehuwd</i>	18+	186,8	6632	6818,3	2,7	2,2	142,7	187,8	244,0
<i>Gescheiden</i>	0- 5	99,0	1890	1989,1	5,0	6,6	76,6	98,9	126,8
<i>Gescheiden</i>	5-18	147,6	3988	4135,5	3,6	5,8	114,8	148,1	189,2
<i>Gescheiden</i>	18+	108,4	3564	3672,1	3,0	5,8	83,9	108,5	137,8
Totaal		2818,9	69342	72160,9	3,9	7,3	2202,4	2827,4	3580,0

Ante	Druggebr.	Niet in GBA	in GBA	Totaal	% niet in GBA	% in HKS	Marge niet in GBA		
							2,5 %	50 %	97,5 %
1	<i>Nee</i>	1454,9	19148	20603,0	7,1	8,4	1110,6	1453,1	1857,9
1	<i>Ja</i>	5,4	236	241,0	2,2	7,7	0,0	0,0	28,1
2+	<i>Nee</i>	1279,8	39147	40427,0	3,2	7,0	952,5	1290,2	1673,9
2+	<i>Ja</i>	78,7	10811	10889,8	0,7	6,4	16,2	75,7	166,8
Totaal		2818,9	69342	72160,9	3,9	7,3	2202,4	2827,4	3580,0

-Tabel 2 vervolg-

2001

Leeftijd		Niet		Totaal	% niet		Marge niet in GBA		
		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
<i>Man</i>	<i>13-17</i>	81,3	3294	3375,3	2,4	13,7	39,2	79,9	132,2
<i>Man</i>	<i>18-24</i>	605,3	6583	7188,3	8,4	15,7	483,3	600,6	737,5
<i>Man</i>	<i>25-44</i>	1145,6	17903	19048,6	6,0	13,8	963,6	1145,6	1322,9
<i>Man</i>	<i>45+</i>	440,4	8725	9165,4	4,8	5,2	276,6	432,5	621,3
<i>Vrouw</i>	<i>13-17</i>	180,7	3236	3416,7	5,3	4,6	75,3	179,2	318,5
<i>Vrouw</i>	<i>18-24</i>	692,4	6822	7514,4	9,2	5,2	474,5	689,3	948,7
<i>Vrouw</i>	<i>25-44</i>	1436,7	17524	18960,7	7,6	3,7	1066,3	1442,2	1846,9
<i>Vrouw</i>	<i>45+</i>	1223,2	10900	12123,2	10,1	1,0	589,5	1214,2	2074,9
Totaal		5805,6	74987	80792,6	7,2	7,5	4939,1	5811,3	6823,7

Verblijfsd.		Niet		Totaal	% niet		Marge niet in GBA		
		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
<i>Ongehuwd</i>	<i>0- 5</i>	2179,3	19907	22086,5	9,9	10,6	1839,3	2181,5	2538,4
<i>Ongehuwd</i>	<i>5-18</i>	1478,9	20467	21945,7	6,7	8,8	1260,8	1481,1	1735,1
<i>Ongehuwd</i>	<i>18+</i>	406,6	7169	7575,6	5,4	8,8	345,5	406,1	483,9
<i>Gehuwd</i>	<i>0- 5</i>	315,5	4010	4325,5	7,3	4,7	267,1	316,0	373,2
<i>Gehuwd</i>	<i>5-18</i>	366,8	6451	6817,9	5,4	2,4	306,2	367,8	442,2
<i>Gehuwd</i>	<i>18+</i>	321,5	6934	7255,6	4,4	2,1	265,5	321,4	392,3
<i>Gescheiden</i>	<i>0- 5</i>	233,6	2127	2360,9	9,9	7,0	197,4	234,0	275,2
<i>Gescheiden</i>	<i>5-18</i>	295,1	4127	4422,1	6,7	5,5	249,9	295,9	348,6
<i>Gescheiden</i>	<i>18+</i>	208,3	3795	4002,9	5,2	5,4	176,1	208,8	248,1
Totaal		5805,6	74987	80792,6	7,2	7,5	4939,1	5811,3	6823,7

Ante		Niet		Totaal	% niet		Marge niet in GBA		
Druggebr.		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
<i>1</i>	<i>Nee</i>	3201,2	20128	23328,7	13,7	9,0	2635,3	3207,7	3797,7
<i>1</i>	<i>Ja</i>	10,7	277	287,9	3,7	8,3	0,0	12,3	38,8
<i>2+</i>	<i>Nee</i>	2485,7	43464	45950,1	5,4	7,1	2038,3	2489,1	3039,2
<i>2+</i>	<i>Ja</i>	108,0	11118	11225,9	1,0	6,4	32,7	104,7	204,7
Totaal		5805,6	74987	80792,6	7,2	7,5	4939,1	5811,3	6823,7

-Tabel 2 vervolg-

2002

Leeftijd		Niet		Totaal	% niet		Marge niet in GBA		
		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
<i>Man</i>	<i>13-17</i>	149,0	3328	3477,0	4,3	15,2	89,9	145,9	208,8
<i>Man</i>	<i>18-24</i>	1098,6	6828	7926,6	13,9	18,1	943,9	1104,8	1272,2
<i>Man</i>	<i>25-44</i>	1934,6	18441	20375,6	9,5	16,1	1712,3	1935,1	2148,9
<i>Man</i>	<i>45+</i>	733,0	9483	10216,0	7,2	6,1	528,0	730,7	978,4
<i>Vrouw</i>	<i>13-17</i>	175,7	3222	3397,7	5,2	4,8	69,0	168,2	308,4
<i>Vrouw</i>	<i>18-24</i>	940,2	7113	8053,2	11,7	5,0	666,5	924,7	1220,3
<i>Vrouw</i>	<i>25-44</i>	2435,6	17929	20364,6	12,0	4,0	1941,9	2439,4	2942,9
<i>Vrouw</i>	<i>45+</i>	863,1	11771	12634,1	6,8	1,4	437,5	857,9	1406,0
Totaal		8329,9	78115	86444,9	9,6	8,6	7549,0	8331,2	9210,2

Verblijfsd.		Niet		Totaal	% niet		Marge niet in GBA		
		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
<i>Ongehuwd</i>	<i>0- 5</i>	2864,4	19855	22719,5	12,6	12,1	2594,6	2860,5	3159,3
<i>Ongehuwd</i>	<i>5-18</i>	2211,0	21878	24089,2	9,2	10,2	1987,3	2213,0	2451,7
<i>Ongehuwd</i>	<i>18+</i>	711,7	7950	8661,3	8,2	10,3	639,7	712,5	785,9
<i>Gehuwd</i>	<i>0- 5</i>	458,0	3961	4419,0	10,4	5,4	408,3	457,6	512,1
<i>Gehuwd</i>	<i>5-18</i>	548,4	6673	7221,3	7,6	2,7	483,2	548,2	617,7
<i>Gehuwd</i>	<i>18+</i>	518,1	7301	7818,7	6,6	2,6	459,6	518,4	581,3
<i>Gescheiden</i>	<i>0- 5</i>	283,1	2065	2348,2	12,1	7,0	252,2	283,1	314,6
<i>Gescheiden</i>	<i>5-18</i>	399,2	4267	4666,1	8,6	5,9	355,9	400,1	446,3
<i>Gescheiden</i>	<i>18+</i>	336,1	4166	4501,6	7,5	6,0	299,3	336,2	374,9
Totaal		8329,9	78115	86444,9	9,6	8,6	7549,0	8331,2	9210,2

Ante		Niet		Totaal	% niet		Marge niet in GBA		
Druggebr.		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
<i>1</i>	<i>nee</i>	4819,1	26501	31320,0	15,4	10,0	4296,4	4814,5	5373,7
<i>1</i>	<i>ja</i>	0,6	102	102,6	0,6	9,0	0,0	0,0	10,7
<i>2+</i>	<i>nee</i>	3402,6	41262	44664,8	7,6	7,9	2953,8	3400,4	3882,1
<i>2+</i>	<i>ja</i>	107,6	10250	10357,7	1,0	7,1	39,6	106,5	185,8
Totaal		8329,9	78115	86444,9	9,6	8,6	7549,0	8331,2	9210,2

-Tabel 2 vervolg-

2003

	Leeftijd	Niet		Totaal	% niet		Marge niet in GBA		
		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
<i>Man</i>	<i>13-17</i>	120,0	3180	3300,0	3,6	15,1	67,9	117,5	176,8
<i>Man</i>	<i>18-24</i>	1466,5	6788	8254,5	17,8	18,1	1279,0	1465,0	1657,7
<i>Man</i>	<i>25-44</i>	1945,3	18414	20359,3	9,6	17,0	1731,4	1944,9	2165,6
<i>Man</i>	<i>45+</i>	743,0	10182	10925,0	6,8	7,3	548,7	733,7	944,3
<i>Vrouw</i>	<i>13-17</i>	152,7	3007	3159,7	4,8	4,3	60,8	148,1	294,0
<i>Vrouw</i>	<i>18-24</i>	1813,1	7102	8915,1	20,3	5,4	1448,1	1823,3	2252,8
<i>Vrouw</i>	<i>25-44</i>	2555,6	17875	20430,6	12,5	4,1	2050,7	2554,0	3147,3
<i>Vrouw</i>	<i>45+</i>	2351,4	12492	14843,4	15,8	1,3	1522,5	2327,6	3379,2
Totaal		11147,6	79040	90187,6	12,4	8,7	10041,2	11120,9	12413,1

	Verblijfsd.	Niet		Totaal	% niet		Marge niet in GBA		
		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
<i>Ongehuwd</i>	<i>0- 5</i>	3159,1	16810	19969,2	15,8	11,9	2843,9	3155,2	3552,1
<i>Ongehuwd</i>	<i>5-18</i>	3375,4	24236	27611,9	12,2	10,6	3029,8	3371,7	3755,6
<i>Ongehuwd</i>	<i>18+</i>	1216,4	9290	10506,7	11,6	10,6	1095,1	1213,4	1356,1
<i>Gehuwd</i>	<i>0- 5</i>	507,8	3151	3658,8	13,9	6,1	453,2	506,8	571,5
<i>Gehuwd</i>	<i>5-18</i>	822,8	6949	7771,9	10,6	2,9	729,8	822,7	931,5
<i>Gehuwd</i>	<i>18+</i>	855,6	7902	8758,1	9,8	3,0	758,2	855,2	970,3
<i>Gescheiden</i>	<i>0- 5</i>	259,1	1630	1889,2	13,7	9,0	230,7	258,3	292,5
<i>Gescheiden</i>	<i>5-18</i>	497,6	4433	4930,1	10,1	5,8	443,4	497,5	564,6
<i>Gescheiden</i>	<i>18+</i>	453,9	4638	5091,7	8,9	5,9	403,6	454,0	513,9
Totaal		11147,6	79040	90187,6	12,4	8,7	10041,2	11120,9	12413,1

Ante	Druggebr.	Niet		Totaal	% niet		Marge niet in GBA		
		in GBA	in GBA		in GBA	in HKS	2,5 %	50 %	97,5 %
1	<i>nee</i>	6450,0	29377	35827,2	18,0	10,0	5777,8	6438,2	7273,8
1	<i>ja</i>	5,7	198	203,8	2,8	9,4	0,0	0,0	24,0
2+	<i>nee</i>	4596,3	40736	45331,8	10,1	8,0	4035,2	4594,6	5263,5
2+	<i>ja</i>	95,7	8729	8824,9	1,1	7,4	29,4	92,5	175,3
Totaal		11147,6	79040	90187,6	12,4	8,7	10041,2	11120,9	12413,1

4. Conclusie

Er is een groei van het aantal Antillianen dat *wel* in de GBA staat ingeschreven van ongeveer 69.000 in 2000 tot 79.000 in 2003. Wij schatten dat het aantal Antillianen dat *niet* in de GBA is geregistreerd van het jaar 2000 tot het jaar 2003 is opgelopen van ongeveer 3.000 tot 12.000. Dit correspondeert met een toename van ongeveer 4 % tot 13 % van de totale groep in Nederland verblijvende Antillianen (dus al dan niet in de GBA ingeschreven).

Het onderzoek kent een aantal beperkingen:

- Het is gezien de wijze van registreren in de HKS en de GBA niet goed mogelijk om een onderscheid aan te brengen tussen Antillianen en Arubanen.
- Het is door de aard van de bestanden de daarin aanwezige gegevens niet mogelijk om bij dit onderzoek een onderscheid aan te brengen tussen verschillende generaties waaruit Antillianen afkomstig zijn. 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 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.

Er zijn bij het maken van de schattingen aannamen gemaakt die uitgebreid beschreven zijn in paragraaf 2. 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 er aannamen zijn geschonden (wat overigens niet bekend is), dit waarschijnlijk heeft geleid tot te hoge schattingen.

Effecten van de schendingen van de aannamen zijn te onderzoeken met behulp van gevoeligheidsanalyses. Men neemt dan een bepaalde schending van de data aan en beziet wat het effect op de schattingen is. Het resultaat is dan een idee van de robuustheid van de schattingen tegen schending van de aannamen. In het kader van deze 'quick scan' hebben wij afgezien van dergelijke gevoeligheidsanalyses.

Een belangrijke aanname is dat, gegeven de gebruikte extra variabelen (leeftijd, geslacht, enzovoort), de kans op opname in HKS en opname in GBA statistisch onafhankelijk van elkaar zijn. In toekomstig onderzoek kan echter wel gepoogd worden deze aanname te omzeilen, door gebruik te maken van drie in plaats van twee registraties. Indien er drie registraties zijn, dan kunnen modellen worden geschat die toestaan dat verschillende kansen op opname aan elkaar gerelateerd zijn.

Een ander onderwerp is de vraag is de aannamen daadwerkelijk zijn geschonden. Dit is niet met behulp van voorliggende data te onderzoeken. Dit is mogelijk te onderzoeken met kwalitatief onderzoek.

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 ‘Integratiekaart’, het bestand aangeleverd door de KLPD onder HKS. Wij hebben de volgende variabelen (aangegeven met hun variabele naam) gebruikt uit beide bestanden:

	Integratiekaart	HKS
Leeftijd	<i>T1</i>	<i>Geb_jaar</i>
Geslacht	<i>v5</i>	<i>Geslacht</i>
Regio	<i>v4</i>	<i>Woon_reg</i>
Verblijfsduur	<i>v11</i>	
Burgerlijke staat	<i>v34</i>	
Som antecedenten		<i>som_ant</i>
Harddruggebruiker		<i>g2</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 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 politieregio's van de vier grote steden (Haaglanden, Rotterdam-Rijnmond, Amsterdam-Amstelland, Utrecht) (ii) de overige politieregio's. Voor de personen die zowel in HKS als in de GBA geregistreerd zijn (voor het jaar 2000 n=4,532), zijn er twee coderingen beschikbaar; deze waren in 167 gevallen niet identiek en dit is opgelost door de codering van HKS te volgen; daarnaast waren er 112 personen waarvan de regio in zowel HKS als GBA onbekend was (inclusief Koninklijke Marechaussee), en deze zijn door ons in de categorie (ii) geplaatst. Deze variabele Regio bleek in de analyses nauwelijks van invloed te zijn en daarom is deze variabele uiteindelijk weggelaten.

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

Verblijfsduur: hier is gekozen voor (i) 0-5 (ii) 5-18 en (iii) 18+.

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

Harddruggebruiker: we merken hier op dat de validiteit van deze variabele niet optimaal is.

Appendix 2: mogelijke koppelingsfouten

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 hier onderzoek naar verricht.

Type (i):

In de GBA hebben mensen een identificatienummer (A-nummer). Bij aanhouding door de politie wordt dit 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 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 validiteitprobleem 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 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.

- 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’ analyse⁴

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

⁴ Deze appendix is een ingekorte en gewijzigde versie van van der Heijden en Hiraseng (2002).

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 hoofdtekst.

In potentie kan een variabele die invloed heeft op de kans van voorkomen in zowel 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 (2004, 2005). 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 (2004 of 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. Hirasig (2002). De vangst-hervangstmethode. In: L.M. Bouter en F.R. Rosendaal. *Dwalingen in de methodologie. Een bundel artikelen uit het Nederlands Tijdschrift voor de Geneeskunde*. Amsterdam: Nederlands Tijdschrift voor geneeskunde. Pp. 84-87.
- Hirasig, 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.
- Van der Vliet, R. (ongedateerd). *Beschrijvingen variabelen van de Integratiekaartbestanden*. Voorburg: CBS.
- Zwane, E., K. van der Pal and P.G.M. van der Heijden (2004). The multiple-record systems estimator when registrations refer to different but overlapping populations. *Statistics in Medicine*, 23, 2267--2281.
- Zwane, E., and P.G.M. van der Heijden (2005). *Analyzing capture-recapture data when some variables of heterogeneous catchability are not observed in all registrations*. Unpublished manuscript.