



OPHOGEN OP PERSOONSNIVEAU VAN GEGEVENS VAN DE LANDELIJKE MEDISCHE REGISTRATIE GEKOPPELD MET DE GBA

Marije van Sijl en Jos de Ree

Samenvatting: In deze nota wordt beschreven hoe bij het maken van statistieken op persoonsniveau gebaseerd op gegevens van aan de GBA gekoppelde registraties met beperkte uniciteit van de koppelsleutel, gebruik gemaakt kan worden van het concept “altijd unieke personen”. Uitgaande van dit concept wordt de ophoging beschreven van persoonsstatistieken op basis van de Landelijke Medische Registratie, een registratie van ziekenhuisopnamen, gekoppeld met de GBA.

Trefwoorden: ophogen, altijd unieke personen, ziekenhuisopnamen, gezondheidsstatistieken, GSB, LMR, GBA.

1. Inleiding

Het CBS maakt bij het produceren van statistieken over zorg en gezondheid in toenemende mate gebruik van externe zorgregistraties. Via koppeling met de Gemeentelijke Basisadministratie (GBA) worden deze samengebracht in een geïntegreerd systeem op persoonsniveau, het zogenaamde GezondheidsStatistisch Bestand (GSB).

De Landelijke Medische Registratie (LMR) van Prismant is de eerste externe registratie die door het CBS in het kader van de ontwikkeling van het GSB aan de GBA is gekoppeld. De LMR is een registratie van ziekenhuisopnamen en bevat

gegevens over uit het ziekenhuis ontslagen patiënten. Dit zijn gegevens op opnameniveau; een patiënt met meerdere opnamen komt meer dan eens in de LMR voor. Naast de administratieve opname- en ontslaggegevens worden in de LMR ook gegevens over diagnoses en verrichtingen vastgelegd. De GBA bevat gegevens over alle in Nederland in het bevolkingsregister ingeschreven personen.

Omdat voor de koppeling van de LMR met de GBA geen volledig identificerende koppelsleutel beschikbaar is, kunnen niet alle LMR-opnamen uniek gekoppeld worden aan personen in de GBA. Koppeling vindt primair plaats op basis van de variabelen geslacht, geboortedatum en postcode (waarvan meestal alleen de vier cijfers beschikbaar zijn). De kans dat een opname uniek gekoppeld kan worden aan een persoon in de GBA is niet voor alle bevolkingsgroepen hetzelfde en kan veranderen in de tijd. Zo hebben personen die in dichtbevolkte postcodegebieden wonen en personen afkomstig uit landen waar de geboortedatum niet altijd exact geregistreerd is, een kleinere kans om uniek gekoppeld te worden. Ook kan de koppelkans veranderen bij verhuizingen. Meer informatie over het koppelen en aspecten die daarbij een rol spelen is te vinden in het methodologisch rapport over de koppeling (Projectgroep ontwikkeling GezondheidsStatistisch Bestand, 2003).

Om op basis van deze niet volledig en enigszins selectief gekoppelde gegevens toch landelijk representatieve statistieken te kunnen maken zijn ophoogmethoden ontwikkeld. Daarbij wordt onderscheid gemaakt tussen ophoging op opnameniveau en ophoging op persoonsniveau.

In een eerdere nota (De Ree & Van Sijl, 2005) is beschreven hoe de uniek gekoppelde LMR-GBA-records worden opgehoogd voor het maken van gezondheidsstatistieken op opnameniveau. Voorbeelden van dit soort statistieken zijn: het gemiddeld aantal opnamen, of de gemiddelde verpleegduur, per persoon in de bevolking per jaar. Deze statistieken kunnen in principe ook alleen op basis van de LMR gemaakt worden, maar door de koppeling met de GBA is het mogelijk om deze te differentiëren naar verschillende bevolkingsgroepen. Door de koppeling van de LMR-opnamen aan de persoonsrecords in de GBA is het namelijk mogelijk om kenmerken uit de GBA (b.v. herkomstgroepering) of uit andere bestanden die aan de GBA gekoppeld zijn (b.v. sociaal-economische gegevens) aan de patiëntgegevens van de LMR toe te voegen.

In deze nota wordt ingegaan op de ophoging van LMR-GBA-statistieken op persoonsniveau, dat wil zeggen statistieken waarvoor het noodzakelijk is om dezelfde persoon gedurende een bepaalde tijdsperiode te volgen. In dit geval worden personen gevolgd op het optreden van ziekenhuisopnamen. Voorbeelden van dit soort persoonsniveaustatistieken zijn: het aantal personen dat een of meer keer in een jaar in het ziekenhuis is opgenomen geweest, of het aantal personen dat voor het eerst in 5 jaar in het ziekenhuis is opgenomen geweest. Deze statistieken kunnen alleen op basis van de gekoppelde LMR-GBA-gegevens gemaakt worden, waarbij bovendien - zoals hierboven ook beschreven voor de statistieken op opnameniveau - door de koppeling met de GBA de mogelijkheid bestaat tot differentiatie naar bevolkingsgroepen. De mogelijkheden voor ophogen zijn echter beperkter voor deze

persoonsniveaustatistieken in vergelijking met die voor de statistieken op opnameniveau, omdat niet goed gecorrigeerd kan worden voor de ontbrekende records in de LMR en voor de records die door onvolkomenheden in de koppelsleutel niet koppelen met de GBA.

Bij de ophoging op persoonsniveau wordt uitgegaan van personen die in de GBA gedurende een bepaalde tijd voortdurend uniek koppelbaar zijn op de beschikbare koppelsleutel (de zogenaamde ‘altijd unieke personen’). De ophoogmethode wordt hier specifiek uitgewerkt voor de koppeling van de LMR aan de GBA, maar kan ook gebruikt worden bij het ophogen van andere registraties met een onvolledig identificerende koppelsleutel. Bij de ophoging op persoonsniveau wordt ten dele ook gebruik gemaakt van methoden of tussenresultaten van de ophoging op opnameniveau; voor details hierover wordt dan verwezen naar de betreffende nota (De Ree & Van Sijl, 2005).

2. Koppeling van LMR met GBA

In de LMR (en ook in veel andere registraties in de gezondheidszorg) is geen uniek landelijk identificatienummer van de patiënt beschikbaar, omdat de registraties van oorsprong anoniem zijn opgezet. Daardoor zijn ook de andere variabelen waarmee gekoppeld kan worden maar beperkt identificerend. Door gebruik te maken van een combinatie van verschillende beschikbare variabelen is het toch mogelijk gebleken om een groot deel van de ziekenhuisopnamen uit de LMR uniek te koppelen aan een persoon in de GBA. Als koppelsleutel worden hierbij de variabelen geslacht, geboortedatum en postcode gebruikt. Bij enkele ziekenhuizen is de volledige postcode beschikbaar, bij de meeste ziekenhuizen echter alleen de cijfers van de postcode. Op basis van de koppelsleutel geboortedatum, geslacht en postcode (4 of 6 posities) kan ca. 87% van de ziekenhuisopnamen uniek gekoppeld worden aan de GBA; als van de postcode alleen gebruik gemaakt wordt van de vier cijfers dan kan ca. 84% uniek gekoppeld worden (zie Projectgroep ontwikkeling GezondheidsStatistisch Bestand, 2003).

Bij het koppelen kunnen drie situaties optreden:

- een LMR-record koppelt aan één GBA-record (unieke koppeling)
- een LMR-record koppelt aan meer dan één GBA-record (meervoudige koppeling)
- een LMR-record koppelt aan geen enkel GBA-record (geen koppeling)

Meervoudige koppeling vindt plaats wanneer er in de GBA meer dan één persoon met dezelfde koppelsleutel staat geregistreerd. Deze zogenaamde ‘administratieve meerlingen’ zijn in dit geval personen van gelijk geslacht, met dezelfde geboortedatum en wonende in hetzelfde postcodegebied. Soms zullen dit echte meerlingen zijn (personen met dezelfde ouders), maar in veel gevallen is er geen sprake van directe verwantschap, vandaar de term ‘administratieve meerlingen’.

Vanwege de beperkte uniciteit van de LMR-GBA-koppelsleutel komen administratieve meerlingen bij deze koppeling redelijk vaak voor (9-12%, afhankelijk van of ook gekoppeld wordt met de postcodes van 6 posities of alleen met de 4-cijferige postcode).

Sommige LMR-records kunnen niet gekoppeld worden aan de GBA (gem. 3,5%). Dit kan verschillende oorzaken hebben, o.a. het niet doorgeven van verhuizingen in het ziekenhuis of bij de gemeente en administratieve fouten in de registraties. Ook kan er soms een verschil zijn in administratieve praktijk: in de GBA worden bij personen met onvolledig bekende geboortedatum soms nullen bij de geboortemaand en geboortedag geregistreerd; bij de LMR komt dit niet voor. In het methodologische rapport over de koppeling worden de niet koppelende records verder beschreven (Projectgroepgroep ontwikkeling GezondheidsStatistisch Bestand, 2003).

Een klein deel van de ziekenhuisopnamen (gemiddeld 0,9%) wordt om verschillende redenen niet geregistreerd in de LMR. Prismant heeft hiervoor, op basis van wel bekende randtotalen, records gegenereerd. Deze records zijn onder een apart ziekenhuisnummer te herkennen in de bestanden. Omdat de gegevens in deze records geen werkelijke personen betreffen zijn deze van te voren uitgesloten van de koppeling aan de GBA.

3. Het concept ‘altijd unieke personen’

Bij het maken van statistieken op persoonsniveau, waarbij personen gedurende een bepaalde periode moeten worden gevolgd op ziekenhuisopnamen, is het belangrijk om voor deze periode alle opnamen van een persoon te kunnen identificeren middels een koppeling aan de GBA. Om dit mogelijk te maken is er voor gekozen om bij de berekening van deze statistieken uit te gaan van alleen die personen van wie vrijwel zeker is dat elke eventuele ziekenhuisopname die in de betreffende periode heeft plaatsgevonden uniek koppelbaar is aan de GBA.

Voor de LMR is dit uitgewerkt door voor alle personen in de GBA na te gaan op welke momenten zij uniek identificeerbaar zijn op de koppelsleutel. Hierbij wordt uitgegaan van de minimale koppelsleutel die voor alle ziekenhuisopnamen beschikbaar is (geslacht, geboortedatum, 4 cijfers van de postcode). Voor alle personen die vanaf 1995 in de GBA ingeschreven staan of hebben gestaan worden deze gegevens verzameld in een zogenaamd ‘uniciteitsbestand’. In dit bestand staan per persoon de tijdsperiodes dat hij/zij wel of niet aanwezig was (gemarkeerd door geboorte, immigratie, sterfte en emigratie) en welke periodes zijn/haar koppelsleutel uniek identificeerbaar was. Per te berekenen uitkomstmaat worden vervolgens hieruit de personen geselecteerd die gedurende de gehele periode waarop de uitkomstmaat betrekking heeft (de onderzoeksperiode), uniek te identificeren zijn. Deze populatie wordt de ‘altijd unieke populatie’ genoemd en is de populatie die gebruikt wordt voor de afleiding van de betreffende persoonsstatistiek. Om deze populatie op te hogen naar de totale Nederlandse bevolking (referentiepopulatie)

worden gewichten berekend. De gebruikte methodiek om deze gewichten te berekenen wordt uitgewerkt in paragraaf 4. Hieronder wordt een algemene afbakening gegeven van de altijd unieke populatie en de referentiepopulatie, welke vervolgens per type uitkomstmaat voor de LMR nader worden gedefinieerd.

Tot de *altijd unieke populatie* behoren de GBA-personen die gedurende de gehele onderzoeksperiode uniek koppelbaar zijn op de koppelsleutel. Daarnaast worden ook de personen die gedurende de onderzoeksperiode zijn geboren of zijn overleden en daarna, respectievelijk daarvoor, de gehele tijd uniek zijn, tot de altijd unieke populatie gerekend. Aangezien personen niet voor hun geboorte of na hun overlijden in het ziekenhuis kunnen worden opgenomen, kan ook van deze personen elke eventuele opname uniek gekoppeld worden.

Tot de *referentiepopulatie* behoren alle personen die gedurende de hele onderzoeksperiode in de GBA ingeschreven hebben gestaan en de personen die tijdens de onderzoeksperiode zijn geboren of zijn overleden. De referentiepopulatie betreft dus een populatie van in principe alle GBA-personen die in de onderzoeksperiode ingeschreven hebben gestaan, ook de niet (altijd) uniek koppelbaren. De personen die gedurende de onderzoeksperiode zijn geïmmigreerd of geëmigreerd worden echter van de referentiepopulatie uitgesloten.

Van de personen die gedurende een deel van de onderzoeksperiode niet in de GBA geregistreerd hebben gestaan vanwege immigratie en/of emigratie is namelijk niet met zekerheid vast te stellen of ze vóór inschrijving dan wel na uitschrijving in een ziekenhuis opgenomen zijn geweest. Daarom worden deze personen buiten beschouwing gelaten voor de uitkomstmaten op persoonsniveau. Ze kunnen niet tot de altijd unieke populatie behoren en ze worden vervolgens ook uitgesloten van de referentiepopulatie.

Per statistiekjaar en per soort uitkomstmaat (afhankelijk van de onderzoeksperiode waarop deze betrekking heeft) zijn er verschillende altijd unieke populaties en referentiepopulaties. De beoogde uitkomstmaten voor de LMR zijn onder te verdelen in drie typen.

Type 1: jaarprevalenties (personen met 1 of meer opnamen in jaar t)

Type 2: incidenties (personen met een eerste opname in jaar t en geen eerdere opnamen in de 5 voorafgaande kalenderjaren)

Type 3: incidenties met volgen op sterfte of heropnamen in jaar t en t+1.

Bij de uitkomstmaten van type 2 en type 3 (incidenties) is er voor de eenvoud voor gekozen om bij het bepalen van het al of niet aanwezig zijn van eerdere opnamen voor iedereen een terugkijkperiode van de 5 voorafgaande kalenderjaren te nemen, in plaats van per persoon 5 jaren terug te kijken uitgaande van de opnamedatum van de eerste opname in jaar t. Bij de uitkomstmaten van type 3 wordt verder voor het volgen op sterfte en heropnamen exact een periode van 1 jaar (12 maanden) na eerste opname genomen, maar de altijd unieke populaties waarvoor dit berekend wordt worden wel – weer voor de eenvoud – geselecteerd op uniciteit gedurende de hele kalenderjaren (jaar t en t+1).

De altijd unieke populatie en de referentiepopulatie van type 2 uitkomstmaten zijn een verbijzondering van de populaties voor type 1; voor de type 3 uitkomstmaten worden de populaties nog verder verbijzonderd. Dit wordt hieronder verder gedefinieerd op basis van de statussen in het GBA-uniciteitsbestand. Deze statussen kunnen als volgt worden samengevat:

- 2: persoon reeds overleden
- 1: persoon nog niet geboren
- 1: persoon uniek op koppelsleutel
- 2: persoon niet uniek op koppelsleutel
- 9: persoon niet ingeschreven in de GBA (is voor immigratie, na emigratie)

Aan de hand hiervan worden de altijd unieke populaties en referentiepopulaties geselecteerd zoals aangegeven in tabel 1.

Tabel 1: Afbakening van de altijd unieke populaties en referentiepopulaties per type uitkomstmaat

	Altijd unieke populatie	Referentiepopulatie
Type 1 Jaar-prevalenties	Gedurende geheel jaar t: Status is 1; of Status is -1 en daarna 1; of Status is 1 en daarna -2; of Status is -1, daarna 1, daarna -2	Gedurende geheel jaar t: Status is nooit 9
Type 2 Incidenties	Zelfde condities voor jaar t als bij type 1 en daarbij geldt ook van jaar t-5 t/m jaar t-1: Status is 1; of Status is altijd -1; of Status is -1, daarna 1	Zelfde condities voor jaar t als bij type 1 en daarbij geldt ook van jaar t-5 t/m t-1: Status is nooit 9 Status is nooit -2
Type 3 Incidenties met volgen in de tijd	Zelfde condities als bij type 1 en 2 en daarbij geldt ook in jaar t+1: Status is 1 of Status is 1 en daarna -2	Zelfde condities als bij type 1 en 2 en daarbij geldt ook in jaar t+1: Status is nooit 9 Status is nooit -1

Bij het volgen in de tijd van personen met een incidentie opname op sterfte (een van de type 3 uitkomstmaten) is het in principe niet nodig om ook in jaar t+1 uniek te zijn, omdat van een eenmaal uniek geïdentificeerd persoon in de GBA altijd de sterfte teruggevonden kan worden. Voor het uitsluiten van de emigranten uit de referentiepopulatie is het echter wel nodig om de status in jaar t+1 mee te nemen. Daarom is voor de uitkomstmaat incidentie met volgen op sterfte toch gekozen om uit te gaan van de populaties volgens type 3.

Bij de toepassing van de ophoogmethode met ‘altijd unieke populaties’ op de LMR, wordt er van uit gegaan dat de registratie van opnamen volledig is en dat alle opnamerecords, tenminste voorzover deze GBA-ingezetenen betreffen, uniek of meervoudig koppelbaar zijn met de GBA. Dit is echter niet helemaal het geval. Zoals in paragraaf 2 al is aangegeven, is de LMR niet geheel volledig (ca. 0,9 %

missende records over 1995-2003) en ca. 3,5% van de records koppelt niet met de GBA (zie tabel 2). Omdat hier op persoonsniveau moeilijk voor te corrigeren valt, betekent dit dat voor met name de statistieken die relateren aan de hele bevolking (zoals de eerder genoemde type 1 en type 2 uitkomstmaten) de berekende cijfers een lichte onderschatting zullen zijn van de werkelijkheid. Deze onderschatting van het aantal personen met opnamen in de bevolking kan niet precies gekwantificeerd worden, maar bedraagt maximaal het totale percentage van missende en niet gekoppelde records (3,5 – 5,6 %). Voor de ophoogmethode van de LMR voor statistieken op opnameniveau (zie De Ree & Van Sijl, 2005) geldt dit probleem niet, omdat hier wel gewogen kan worden naar alle LMR-records, inclusief de niet gekoppelde records en de totalen van de missende records.

De onderschatting van de cijfers van de statistieken op persoonsniveau kan nog enigszins variëren naar bijvoorbeeld leeftijd en diagnose. Zo komen niet gekoppelde LMR-records vaker voor bij jongvolwassenen en zeer ouderen en bij psychische stoornissen en ongevalsletsels. Voor deze diagnoses en leeftijdsgroepen kan de onderschatting van het aantal personen met opnamen dus iets groter zijn. Bij de ontbrekende records in de LMR horen ook de opnamen van één (vanaf 1998) tot twee (van 1995 tot 1997) gespecialiseerde kankerklinieken. Voor diagnoses in de groep nieuwvormingen kan er daardoor ook sprake zijn van een iets grotere onderschatting in de cijfers. Omdat van het totaal aantal opnamen voor nieuwvormingen echter maar een relatief klein gedeelte in deze kankerklinieken plaatsvindt (ca. 4% per kliniek) en een deel van de in deze gespecialiseerde klinieken opgenomen personen al eerder in een ander ziekenhuis zal zijn opgenomen, zal deze onderschatting toch beperkt zijn.

Tabel 2: Percentage missende en niet koppelende records per jaar in de LMR

Jaar	Missende records	Niet koppelende records	Totaal missende en niet koppelende records
	%	%	%
1995	1,49	4,06	5,55
1996	1,38	3,82	5,20
1997	0,96	3,67	4,63
1998	0,94	3,63	4,58
1999	0,64	3,47	4,11
2000	0,90	3,38	4,29
2001	0,88	3,26	4,14
2002	0,47	3,08	3,55
2003	0,57	2,94	3,51

4. Ophoogprocedure

Voor de productie van statistieken op persoonsniveau die representatief moeten zijn voor de totale GBA-populatie wordt de altijd unieke populatie opgehoogd naar de referentiepopulatie. De eenvoudigste manier om dit te doen is door alle altijd unieke personen een gewicht mee te geven dat gelijk is aan de verhouding tussen de omvang van de referentiepopulatie en die van de altijd unieke populatie. Het is echter bekend dat de altijd unieke populatie op bepaalde kenmerken afwijkt van de referentiepopulatie. Zo hangt de kans om altijd uniek te zijn bijvoorbeeld af van leeftijd, geboorteland en postcodegebied. Als men in de totaalcijfers deze verschillende bevolkingsgroepen goed wil representeren is het nodig om de altijd unieke populatie en referentiepopulatie naar de betreffende kenmerken te stratificeren. Door vervolgens per stratum (elke combinatie van waarden van de variabelen in de stratificatie) gewichten te bepalen wordt een betere weergave van de werkelijkheid bereikt.

In de hier uitgewerkte ophoogprocedure zijn de variabelen leeftijd, geslacht, herkomstgroep, corop (regio-indeling), geboortelandgroep en koppelkans meegenomen bij het indelen van de bevolking in strata. Koppelkans betreft de kans van een individu om uniek te zijn op de koppelsleutel (zie De Ree & Van Sijl, 2005). De variabelen zijn ingedeeld in klassen om het aantal verschillende waarden beperkt te houden. In bijlage 1 is de klassenindeling gegeven en waar nodig een toelichting op de variabelen.

Wanneer de populaties volgens slechts één stratificatie (kruising van variabelen) in strata worden ingedeeld, worden de gewichten bepaald door per stratum de verhouding te bepalen tussen de omvang van de referentiepopulatie en de omvang van de op te hogen altijd unieke populatie. Per stratum is de som van de gewichten gelijk aan de omvang van de referentiepopulatie in dat stratum. Bij volledige kruising van alle bovengenoemde variabelen blijken er echter strata te ontstaan waar zeer weinig personen in zitten. Wanneer er in een stratum geen altijd unieke personen zijn, maar wel personen in de referentiepopulatie kan er voor dat stratum niet goed opgehoogd worden. Daarom zijn de strata wat ingedikt en is er voor gekozen om te wegen naar drie verschillende stratificaties, met per stratificatie een beperkt aantal variabelen:

- leeftijd x geslacht x herkomstgroep
- leeftijd x geslacht x regio (corop)
- geboortelandgroep x koppelkansklasse.

De eerste twee kruisingen zijn gekozen omdat dit uitsplitsingen zijn waar naar gepubliceerd zal gaan worden en die daarom zeker goed gemaakt moeten worden bij het ophogen. De laatste kruising betreft variabelen die van grote invloed zijn op het feit of mensen gedurende langere tijd uniek zijn. Bij de variabele ‘koppelkansklasse’ wordt gebruik gemaakt van tussenresultaten van de ophoging van LMR-gegevens op opnameniveau (De Ree & Van Sijl, 2005); in Bijlage 1 wordt de afleiding van deze variabele nader toegelicht.

Voor het wegen naar de verschillende stratumindelingen (stratificaties) wordt gebruik gemaakt van de methode Iterative Proportional Fitting (IPF), ook bekend onder de namen 'Raking Ratio estimation' en 'Rassen' (Deville et al, 1993). Bij deze methode krijgen alle personen uit de populatie van altijd unieken een (positief) gewicht zodanig dat de som der gewichten voor elk stratum van elke stratumindeling gelijk is aan het aantal personen in dat stratum in de referentiepopulatie. De gewichten worden in een aantal iteratieronden bepaald. Eerst worden er volgens de eerste stratumindeling gewichten bepaald zodanig dat de som van de gewichten van de altijd unieken per stratum gelijk is aan de omvang van de referentiepopulatie per stratum. Vervolgens worden de personen in strata ingedeeld volgens de tweede indeling, waarbij de altijd unieke personen al over een gewicht beschikken. De gewichten worden nu weer aangepast zodanig dat de som van de gewichten van de altijd unieken per stratum van de tweede indeling gelijk is aan de omvang van de referentiepopulatie per stratum. Daarna volgt een aanpassing van de gewichten om de derde indeling kloppend te maken. De gewichten die nu na deze eerste weegronde bereikt zijn geven echter weer minder goede resultaten voor de eerste stratumindeling. Daarom worden de gewichten nu opnieuw aangepast om ze voor de eerste indeling kloppend te maken, enz. Wanneer de weegronde een aantal keren herhaald is (iteraties) zijn de gewichten zodanig aangepast dat ze voor alle indelingen een goede ophoging geven. Dit zijn de eindgewichten die gebruikt zullen worden voor het ophogen van de uitkomstmaten op persoonsniveau.

Bij de persoonsniveaugewichten die zijn geproduceerd voor de LMR-uitkomstmaten is gekozen voor 30 iteraties. Na 30 iteratieslagen bleken de gewichten namelijk voldoende te convergeren: de relatieve afwijkingen van de gewogen aantallen altijd unieke personen per stratum en de bijbehorende aantallen in de referentiepopulatie bedroegen toen minder dan 0.001 %. Het aantal iteraties kan overigens in de in de programmatuur eenvoudig naar behoefte bijgesteld worden. Ook kan de programmatuur gebruikt worden voor het ophogen van andere (GBA-)populaties.

Voor de LMR zijn per statistiekjaar en per type uitkomstmaat, met bijbehorende altijd unieke en referentiepopulaties (zie paragraaf 2), aparte ophooggewichten geproduceerd volgens de hier beschreven methode. Door deze gewichten vervolgens toe te passen bij de berekening van de uitkomstmaten wordt de ophoging van de betreffende persoonsstatistieken gerealiseerd.

5. Literatuur

De Ree, Jos & Van Sijl, Marije (2005). Ophogen op opnameniveau van gegevens van de Landelijke Medische Registratie gekoppeld met de GBA. Interne CBS-nota, nr. 0159-05-SOO, 2 mei 2005. CBS, Voorburg.

Deville, J.C., Särndal, C.E. & Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, Vol. **88**, No. 423, pp. 1013-1020.

Projectgroep ontwikkeling GezondheidsStatistisch Bestand (2003). Koppeling van LMR- en GBA-gegevens: methode, resultaten en kwaliteitsonderzoek. (auteurs: A. de Bruin, E.I. de Bruin, A. Gast, J.W.P.F. Kardaun, M. van Sijl, G.C.G. Verweij). CBS, Voorburg/Heerlen.

Bijlage 1

Beschrijving van bij de weging gebruikte variabelen

Leeftijd:

De leeftijd betreft de leeftijd op 31 december van het betreffende statistiekjaar en is ingedeeld in de volgende klassen:

0 jaar

1-4 jaar

5-9 jaar

10-14 jaar

15-19 jaar

etc., t/m

95 jaar en ouder

Geslacht:

Man/vrouw

Herkomstgroepering:

Volgens de CBS-definitie; de volgende groepen zijn onderscheiden:

Autochtonen

Marokko

Turkije

Suriname

Nederlandse Antillen en Aruba

Overige niet-westerse allochtonen

Westerse allochtonen

Corop:

Regionale indeling van Nederland in 40 gebieden

Geboortelandgroep:

Groepering van geboortelands op basis van het patroon van geregistreerde verjaardagen. Personen uit landen met een minder goede bevolkingsadministratie van wie de geboortedatum niet exact bekend is staan in de GBA vaak met dezelfde

geboortedatum (bijv. 1 januari, 1 juli) ingeschreven. Hierdoor wordt de kans om uniek koppelbaar te zijn voor deze personen aanzienlijk verkleind. Voor meer informatie, zie ook De Ree & Van Sijl, 2005.

Code groep	Code geboortelanden	Namen Geboortelanden
01	6030	Nederland
02	6043	Turkije
03	5022	Marokko
04	5043	Irak
05	6023	Afghanistan
06	5012	Iran
07	6013, 9027 en 9028	Somalië
08	5007, 5095, 6024, 7011, 7058, 9030, 5003, 5009, 6029, 7073, 7085, 9089, 5002, 5010, 5017, 6039, 7050, 5001, 5014, 6008, 6014, 6016, 7027, 5027, 7033, 7035, 7036, 9052	Suriname, de Antillen, Nederlands Indië, Nieuw Guinea, Indonesië, Zwitserland, Oostenrijk, Duitsland, Frankrijk, België, Groot-Brittannië, Hongarije, Portugal, Canada, Zuid-Afrika, Brazilië, VS, Australië, Dominicaanse Republiek, Filippijnen, Sri Lanka, Japan, Hongkong, Ceylon
00	Rest	

Koppelkansklasse:

De kans om uniek te koppelen op de koppelsleutel geboortedatum, geslacht en viercijferige postcode hangt onder andere af van het aantal inwoners van het postcode gebied, leeftijd, geslacht en geboorteland. Omdat in de eerste twee stratificaties van de weging op persoonsniveau (zie hoofdstuk 3) de grootte en bevolkingssamenstelling van de afzonderlijke postcodegebieden, die een grote invloed hebben op de koppelkans, nog niet is verdisconteerd, is gekozen voor een derde stratificatie waarin deze koppelkans bepalende factoren wel specifiek worden meegenomen.

Bij de ophoging op opnameniveau zijn al de koppelkansen berekend per viercijferige postcode, geslacht, geboortjaar en geboortelandgroep op 1 januari en 31 december van ieder jaar (zie paragraaf 3.6 van De Ree & Van Sijl, 2005). Op basis van deze gegevens is aan alle personen in de GBA een koppelkans toegekend. Per postcode is vervolgens de gemiddelde koppelkans van alle inwoners uitgerekend op 1 januari en 31 december van elk jaar. Omdat voor de weging alleen variabelen met een beperkt aantal klassen gebruikt kunnen worden, zijn deze gemiddelde koppelkansen per postcode vervolgens ingedeeld in 7 klassen:

1. $\leq 0,70$
2. $> 0,70 \leq 0,75$
3. $> 0,75 \leq 0,80$
4. $> 0,80 \leq 0,85$
5. $> 0,85 \leq 0,90$
6. $> 0,90 \leq 0,95$
7. $> 0,95$

Van iedere persoon wordt op basis van zijn of haar postcode vastgesteld in welke klasse hij of zij aan het begin (= 1 januari van het beginjaar) en het einde (= 31 december van het eindjaar) van de onderzoeksperiode van de betreffende uitkomstmaat valt. De laagste kansklasse wordt ten slotte gebruikt als de koppelkansvariabele in de (derde) stratificatie. De laagste klasse wordt gebruikt omdat de kans om gedurende langere tijd uniek te zijn nooit hoger kan zijn dan de kans op enig moment in de tijd.

Naast de variabele koppelkansklasse, welke dus is bepaald per postcode gebied, wordt in de derde stratificatie ook nog gekruist met de variabele geboortelandgroep. Dit is gedaan om deze belangrijke koppelkans bepalende variabele ook op individueel niveau in de weging gerepresenteerd te hebben.

De mogelijkheid is overwogen om voor de kans op uniciteit van een persoon een model te ontwikkelen waarin alle veranderingen in de tijd worden meegenomen. Vanwege de complexiteit en de benodigde tijdsinvestering om een dergelijk model te ontwikkelen is hier voorlopig niet voor gekozen.