



Centraal Bureau voor de Statistiek

Divisie Sociale en Ruimtelijke Statistieken
Sector Ontwikkeling en Ondersteuning

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

Jos de Ree & Marije van Sijl

Samenvatting: Gegevens over ziekenhuisopnamen uit de Landelijke Medische Registratie (LMR) zijn gekoppeld aan de GBA. Het ophogen van deze gekoppelde gegevens op opnameniveau is uitgevoerd in drie onderdelen: ophoging voor koppelkansen, ophoging voor ontbrekende LMR-records en een finale herweging. In deze nota worden de verschillende onderdelen beschreven.

Trefwoorden: ophogen, 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

Projectnummer: 200021
Archiefnummer: 018-08-SAV
Datum: 14 februari 2008

alle in Nederland in het bevolkingsregister ingeschreven personen. Koppeling van de LMR met de GBA vindt plaats op basis van de variabelen geslacht, geboortedatum en postcode (4 of 6 posities). Voor personen die in het ziekenhuis zijn overleden wordt bovendien gebruik gemaakt van de overlijdensdatum. De details van de gebruikte registraties en de koppelingsmethode zijn elders beschreven (Projectgroep ontwikkeling GezondheidsStatistisch Bestand, 2003).

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. De kans dat een opname uniek gekoppeld kan worden aan een persoon in de GBA is ook 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 en sterfte. Om op basis van de gekoppelde gegevens toch landelijk representatieve statistieken te kunnen maken zijn daarom ophoogmethoden ontwikkeld.

In deze nota wordt 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.

Voor statistieken op persoonsniveau, dat wil zeggen statistieken waarvoor het noodzakelijk is om dezelfde persoon gedurende een bepaalde tijdsperiode te volgen, is een andere ophoogmethode ontwikkeld. 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 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. De details van de ophoging op persoonsniveau zijn in een aparte nota beschreven (Van Sijl & De Ree, 2005).

2. Ophogen op opnameniveau

Voor de ophoging van de gekoppelde LMR-GBA-gegevens op opnameniveau wordt uitgegaan van de ‘maximale’ koppelingsresultaten, d.w.z. de resultaten na koppeling op geboortedatum, geslacht, postcode (eerst op volledige postcode van 6 posities indien deze beschikbaar is, daarna op de 4 cijfers van de postcode) en in geval van overlijden in het ziekenhuis indien nodig ook op de overlijdensdatum. De bij deze ophoging gebruikte gegevens betreffen de koppelingsresultaten na stap III, zoals beschreven in § 4.3 – 4.5 van de publicatie ‘Koppeling van LMR- en GBA-gegevens: methode, resultaten en kwaliteitsonderzoek (Projectgroepgroep ontwikkeling GezondheidsStatistisch Bestand, 2003). Op basis van deze sleutel kan ca. 87% van de ziekenhuisopnamen uniek gekoppeld worden aan de GBA.

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%).

LMR-records kunnen alleen aan de GBA gekoppeld worden als een persoon in beide registraties een geldige koppelsleutel heeft. In de GBA staan personen geregistreerd waarvoor bij de geboortemaand en de geboortedag nullen zijn ingevuld. Deze gegevens leveren geen geldige koppelsleutel op. In de records van de LMR komen deze waarden niet voor, zodat in deze gevallen geen (correcte) koppeling kan plaatsvinden.

Er zijn ook andere redenen waarom LMR-records niet gekoppeld kunnen worden aan de GBA, zoals het niet doorgeven van verhuizingen in het ziekenhuis of bij de gemeente en administratieve fouten in de registraties. In het methodologische rapport over de koppeling worden de niet koppelende records (gem. 3,5%) verder beschreven (Projectgroepgroep ontwikkeling GezondheidsStatistisch Bestand, 2003).

Een klein deel van de ziekenhuisopnamen (aflopend van 1,5 % in 1995 tot 0,5 % in 2002) 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.

Om de uniek gekoppelde LMR-opnamen representatief te maken voor alle LMR-opnamen is een ophoogmethode ontwikkeld. Deze ophoogmethode op opnameniveau valt uiteen in drie onderdelen. Allereerst is er een factor ontwikkeld waarmee de uniek gekoppelde LMR-GBA-records worden opgehoogd om te corrigeren voor de kans op meervoudige koppeling en het hebben van een ongeldige koppelsleutel. Daaraan wordt een factor toegevoegd die corrigeert voor het feit dat sommige ziekenhuisopnamen niet in de LMR geregistreerd worden. Ten slotte vindt er nog een finale herweging plaats om er voor te zorgen dat de ophoogde aantallen uniek gekoppelde opnamen ingedeeld naar een aantal kenmerken overeenkomen met het totaal aantal opnamen volgens de gegevens van Prismant. Hiermee worden ook de afwijkingen ten gevolge van gemiste koppelingen door een niet overeenkomende adressering in GBA en LMR rechtgetrokken.

De drie onderdelen van het ophogen worden hieronder in afzonderlijke hoofdstukken besproken. Er zijn gekoppelde LMR-GBA-jaarbestanden beschikbaar vanaf 1995. De ophoogfactoren worden voor elk kalenderjaar berekend en toegevoegd aan het betreffende jaarbestand van de gekoppelde LMR-GBA-gegevens. In Bijlage 3 wordt ten slotte en overzicht gegeven van alle stappen die in de ophoging zijn uitgevoerd.

3. Berekening van de ophoogfactor voor de koppelkans

De ophoogfactor voor de koppelkans wordt per jaar bepaald, op basis van de GBA-bestanden. Hiervoor worden de volgende berekeningen gemaakt:

- Allereerst worden de kansen berekend op het uniek koppelen van personen met een geldige waarde op de koppelsleutel. Deze kansen worden berekend per geboortjaar, geslacht, postcode (4 cijfers) en geboortelandgroep voor de stand van het begin en eind van elk kalenderjaar. De berekening van de deelfactor 'koppelkans' staat verder uitgewerkt in §3.3. De indeling in geboortelandgroepen staat vermeld in Bijlage 1.
- Vervolgens worden de fracties personen met een geldige koppelsleutel bepaald. De berekening van deze fracties vindt plaats per geslacht, geboortelandgroep en klasse van vijf geboortejaren (groepen 0-4 jaar, 5-9 jaar, 10-14 jaar,, 85 jaar en ouder; op basis van de leeftijd op 31 december) voor de stand van het begin en eind van elk kalenderjaar. De berekening van deze deelfactor ('kansloos' genoemd) staat uitgewerkt in §3.4.

Nadat deze berekeningen zijn uitgevoerd kan een ophoogfactor worden bepaald, zijnde de reciproque waarde van het product van de kans op een unieke koppeling en de fractie met een geldige koppelsleutel. Bij een controle, waarbij de uniek te koppelen personen opgehoogd met deze ophoogfactor werden vergeleken met de populatiegegevens, bleek dat personen tussen 0 en 24 jaar waren

ondervertegenwoordigd en vooral de jongsten. Een belangrijke verklaring voor dit verschijnsel is dat meerlingen van hetzelfde geslacht die op hetzelfde adres wonen niet uniek koppelen. Zodra de kinderen het huis uitgaan verdwijnt dit verschijnsel. Om voor de ondervertegenwoordiging van de jongeren te corrigeren is daarom een extra ophoogfactor bepaald:

- Op het eind van elk kalenderjaar wordt per geboortjaar en geslacht de verhouding bepaald tussen het op basis van de koppelkans en de fractie kanslozen opgehoogde aantal uniek te koppelen personen en het werkelijke aantal personen. Deze deelfactor ('tweeling' genoemd) is uitgewerkt in §3.5.

Op basis van de deelfactoren 'koppelkans', 'kansloos' en 'tweeling' wordt ten slotte de ophoogfactor voor de koppelkans berekend. Dit wordt gedaan door voor het begin en het eind van elk kalenderjaar de deelfactoren te vermenigvuldigen, van het product de reciproque waarde te nemen en tot slot de resulterende ophoogfactoren voor het begin en eind van het kalenderjaar te middelen (zie §3.6). Dit is het eerste onderdeel van de uiteindelijke ophoogfactor.

3.1 Voorbereiding

Eerst wordt uit de GBA een bestand afgeleid waaruit alle voor dit onderwerp overbodige variabelen verwijderd zijn. In stap 1 worden aan dit GBA-bestand twee extra velden toegevoegd: de geboortelandgroep en het dagnummer.

De geboortelandgroep wordt bepaald aan de hand van het geboorteland van de persoon. De indeling in groepen wordt bepaald op basis van het patroon van geregistreeerde geboortedata. Personen uit landen met een minder goede bevolkingsadministratie van wie de geboortedatum niet exact bekend is staan in de GBA vaak met dezelfde geboortedatum ingeschreven (bijv. 1 januari of 1 juli). Hierdoor wordt de kans om uniek koppelbaar te zijn voor deze personen aanzienlijk verkleind. De indeling in geboortelandgroep is te vinden in Bijlage 1.

Het dagnummer loopt op met de dagen van het jaar. Personen zonder geldige geboortedag krijgen dagnummer 0. Dagnummer 1 staat voor 1 januari en dagnummer 366 voor 31 december. Dagnummer 60 staat voor 29 februari en is een geldig dagnummer dat alleen in schrikkeljaren voorkomt.

Het nieuwe bestand wordt gesorteerd op geslacht, geboortjaar, postcode en geboortelandgroep.

3.2 Afleiding hulpbestanden

In stap 2 worden op basis van de records die in stap 1 zijn gecreëerd en die voldoen aan een opgegeven peildatum twee hulpbestanden gegenereerd: *Postcode* en *Groep*. *Postcode* bevat per postcode (4 cijfers), geslacht en geboortjaar het totale aantal personen en de aantallen personen per geboortelandgroep. *Groep* bevat per geboortelandgroep, geslacht en geboortjaar het totale aantal personen met een geldige geboortedag, het aantal personen met een ongeldige geboortedag en de aantallen personen per geldige geboortedag.

3.3 Bepaling koppelkansen

In stap 3 worden per combinatie van postcode (bedoeld wordt steeds de viercijferige postcode), geboorteaar, geslacht en geboortelandgroep koppelkansen berekend. Bij de berekening van de koppelkans wordt rekening gehouden met de ongelijke verdeling van de geboortedagen over het jaar. Daarbij wordt gebruik gemaakt van de gegevens die per postcode en per geboortelandgroep in stap 2 zijn bepaald.

Per geslacht en geboorteaar zijn per postcode het totale aantal inwoners n en het aantal inwoners $m[g]$ per geboortelandgroep g en per geboortelandgroep j het totale aantal inwoners $t[j]$ en het aantal inwoners $a[j,d]$ per geboortedag d gegeven.

Laat $p[j,d]$ de kans zijn dat een inwoner uit geboortelandgroep j op dag d is geboren. De kans dat een persoon geboren op dag d uniek koppelt is dan gelijk aan product van de $(n-1)$ kansen $(1-p[j,d])$ dat de overige inwoners niet op die dag geboren zijn. Dan geldt voor een willekeurige persoon uit geboortelandgroep g van een bepaald geslacht, postcode en geboorteaar dat de kans $p(\text{uniek},g)$ om uniek te koppelen gelijk is aan de som over de dagen d van het jaar van de kans om op dag d geboren te zijn maal de kans dat de $n-1$ overige inwoners niet op deze dag geboren zijn. Dit levert de volgende formule voor de koppelkans op:

$$P(\text{uniek}, g) = \sum_d (p[g,d] \times (1 - p[g,d])^{(m[g]-1)} \times \prod_{j \neq g} (1 - p[j,d])^{m[j]})$$

De kans $p[j,d]$ voor een inwoner uit geboortelandgroep j om op dag d te zijn geboren wordt gebaseerd op de geboortedagverdeling van alle inwoners van Nederland uit die geboortelandgroep. De kansen voor de verschillende dagen worden daarbij bepaald door de aantallen $a[j,d]$ te delen door het totale aantal inwoners $t[j]$.

Indien wordt uitgegaan van een uniforme verdeling van de geboortedagen over het jaar dan is $p[j,d]$ gelijk aan $1/365$ voor een gewoon jaar en $1/366$ voor een schrikkeljaar. Voor een gewoon jaar gaat de formule voor $P(\text{uniek},g)$ dan over in

$$P(\text{uniek}, g) = P(\text{uniek}) = (1 - \frac{1}{365})^{n-1}$$

In de meeste gevallen verschillen de uitkomsten van beide berekeningen niet erg. Voor een aantal geboortelands zijn de verschillen wat groter. Voor deze landen zijn de geregistreerde geboortedagen verre van uniform verdeeld. Dit hangt samen met het feit dat de werkelijke geboortedatum niet bekend is en in plaats daarvan een speciale feestdag of de eerste dag van het jaar of de maand van onze kalender of van de kalender van het geboorteland gekozen wordt.

Daarom is besloten om uit te gaan van de eerste formule, waarbij de kans op een bepaalde geboortedag voor heel Nederland bepaald wordt en alleen afhangt van het geslacht, het geboorteaar en de geboortelandgroep. De deelfactor 'koppelkans' hangt dus af van de kans op een bepaalde geboortedag en de samenstelling van de bevolking van een bepaalde postcode naar geslacht, geboorteaar en geboortelandgroep.

3.4 Bepaling kanslozen

Bij de bepaling van de koppelkansen is ervan uitgegaan dat voor elke inwoner van Nederland een geldige geboortedatum is geregistreerd. Inspectie van het hulpbestand *Groep* leert dat voor een aantal geboortelandgroepen dit niet het geval is. Het gaat hier voornamelijk om personen met geboortedag '00', meestal in combinatie met geboortemaand '00'. Deze personen kunnen niet aan het juiste record gekoppeld worden omdat de LMR in deze gevallen wel een geldige geboortedatum kent.

Om hiervoor te corrigeren wordt in stap 4 per geslacht, geboortjaar en geboortelandgroep bepaald welke fractie van het aantal personen een geldige geboortedatum heeft. Hiertoe worden in een eerste doorloop van het bestand *Groep* per geslacht, geboortelandgroep en geboorteaarklasse de totale aantallen personen met een geldige geboortedag en de aantallen personen zonder geldige geboortedag bepaald en opgeslagen in de arrays *Aantal* en *Totaal*. In een tweede doorloop worden hieruit de fractie personen met een geldige geboortedag afgeleid. Deze fractie worden per geslacht, geboortjaar en geboortelandgroep bepaald en opgeslagen in het bestand *Kansloos*. Deze deelfactor wordt 'kansloos' genoemd, omdat deze bij de ophoging feitelijk corrigeert voor de fractie kanslozen.

3.5 Bepaling deelfactor 'tweeling'

Bij de bepaling van de koppelkansen is het uitgangspunt geweest dat de gebeurtenis dat iemand met een gegeven geslacht, geboortjaar en postcode op een bepaalde dag geboren is onafhankelijk is van de gebeurtenis dat een ander met dezelfde kenmerken op diezelfde dag geboren is. In het geval van tweelingen (en meerlingen) van hetzelfde geslacht geldt dit niet. Zolang zij binnen dezelfde postcode wonen blijft een unieke koppeling uit. Naarmate zij ouder worden zullen zij minder binnen dezelfde postcode blijven wonen en zal dit verschijnsel verdwijnen.

Het zou veel werk kosten en er zou gebruik gemaakt moeten worden van andere gegevens om een echte tweelingfactor te berekenen om de tweelingen van ongelijk geslacht extra te kunnen ophogen ter compensatie voor het ontbreken van een unieke koppeling bij tweelingen van hetzelfde geslacht. Daarvan is voorlopig afgezien. In plaats daarvan is gekozen voor een opzet waarbij per geboortjaar en geslacht de verhouding is bepaald tussen het op basis van de deelfactor *koppelkans* en de deelfactor 'kansloos' opgehoogde aantal uniek te koppelen personen en het werkelijke aantal personen. Deze fractie is voor nuljarigen circa 0,98 en loopt langzaam op tot waarden rond de 1 voor personen vanaf 20 jaar. Deze fractie, de deelfactor 'tweeling' corrigeert voor de ondervetegenwoordiging van jongeren.

De bepaling loopt via stap 5 waarin na een sortering van het totaalbestand op geslacht, geboortedatum en postcode bestanden met de stand op het einde van elk kalenderjaar zijn afgesplitst. In stap 6 wordt vervolgens voor elk eindejaarbestand bepaald wie uniek te koppelen is en wie niet.

In stap 7 worden na sortering op geslacht, geboortjaar, geboortelandgroep en postcode de in stap 4 berekende koppelkansen en fracties kanslozen aangekoppeld

en in stap 8 worden ophooggewichten berekend. Tot slot worden in stap 9 de opgehoogde aantallen berekend en gedeeld door de werkelijke aantallen, hetgeen uiteindelijk leidt tot het bestand *Tweeling*.

3.6 Bepaling van de ophoogfactor koppelkans

De deelfactoren voor de koppelkansen, kanslozen en tweelingen vormen samen het eerste gedeelte van de uiteindelijke ophoogfactor. Deze eerste ophoogfactor, '*Ophoogfactor koppelkans*' genoemd, wordt volgens onderstaande systematiek berekend.

De deelfactoren die een rol spelen bij de ophoging zouden strikt genomen voor elk moment opnieuw moeten worden bepaald. Uit praktische overwegingen wordt hiervan afgezien. De eerste deelfactor *koppelkans*, welke de kans op een unieke koppeling weergeeft, wordt twee keer per jaar bepaald, aan het begin en aan het eind. Ook de tweede deelfactor *kansloos*, welke feitelijk de fractie geldige koppelsleutels weergeeft, wordt voor het begin en het einde van het jaar bepaald. De derde deelfactor *tweeling* wordt slechts eenmaal per jaar berekend, namelijk alleen voor het einde van het jaar. Hierbij wordt verondersteld dat de mate waarin de ondervertegenwoordiging van jongeren voorkomt over het jaar weinig varieert en dat deze factor ook voor het begin van het jaar gebruikt mag worden.

In gevallen waarbij een van de deelfactoren bij gebrek aan gegevens niet kan worden berekend, krijgt de betreffende deelfactor de waarde 1,00. Situaties waarbij dit kan optreden zijn de volgende:

- Voor de oude geboortejaren zijn er soms wel mensen in een postcode op 1 januari en niet op 31 december, omdat alle personen van dat geboortejahr intussen zijn overleden.
- Voor het nieuwste geboortejahr zijn er nog geen personen op 1 januari en wel op 31 december, namelijk de kinderen die in dat jaar geboren zijn.
- Postcodes die ontstaan gedurende een jaar hebben geen kans op 1 januari.
- Postcodes die worden opgeheven in een jaar hebben geen kans op 31 december.
- Personen die tijdens hun ziekenhuisopname in een postcode wonen waar op 1 januari en/of op 31 december geen enkel persoon van het betreffende geboortejahr en geslacht woont.

De reden dat in deze situaties kans 1,00 wordt ingevuld is dat er gedurende enig moment in het jaar 1 persoon van dat geslacht en geboortejahr in die postcode heeft gewoond.

Daarnaast wordt ook voor personen die gekoppeld zijn op basis van de volledige postcode of met het extra gegeven overlijdensdatum, de deelfactor *koppelkans* op 1,00 gezet. De kans om uniek te koppelen op deze uitgebreidere koppelsleutel is namelijk vrijwel gelijk aan 1.

Voor het begin en voor het eind van het jaar kan nu de ophoogfactor voor de koppelkans worden vastgesteld door de drie deelfactoren te vermenigvuldigen en van het product de reciproque waarde te nemen. Tot slot worden de twee ophoogfactoren gemiddeld zodat de opgehoogde uitkomsten de gemiddelde omvang van de populatie weergeven. Dit is de uiteindelijke *Ophoogfactor koppelkans*.

In twee stappen wordt vervolgens deze ophoogfactor aan de LMR-records toegevoegd. Stap 10 splitst het LMR-bestand in jaarbestanden en stap 11 voegt de *Ophoogfactor koppelkans* aan de records van de LMR-bestanden toe.

Daarnaast worden voor andere doeleinden de samengestelde koppelkansen (*koppelkans*kansloos*tweeling*) ook nog per postcode, geslacht, geboortjaar en geboortelandgroep voor het begin en einde van elk jaar in stap 12 in een apart bestand weggeschreven. Dit geaggregeerde bestand met koppelkansen voor het begin en einde van elk jaar wordt namelijk gebruikt bij het ophogen op persoonsniveau (Van Sijl & De Ree, 2005).

4. Berekening van de ophoogfactor voor ontbrekende records

Niet alle ziekenhuisopnamen worden in de LMR geregistreerd: ca. 1% van de opnamen ontbreekt, zie ook § 3.4 van de publicatie ‘Koppeling van LMR- en GBA-gegevens: methode, resultaten en kwaliteitsonderzoek (Projectgroepgroep ontwikkeling GezondheidsStatistisch Bestand, 2003). Prismant genereert hiervoor duplicaatrecords op basis van informatie die wel bekend is, namelijk het totaal aantal opnamen per ziekenhuis naar zorgtype, specialisme en postcode. In een enkel geval zijn de totalen naar postcode niet bekend, maar weet men alleen de gemeente. Prismant kopieert bestaande records met het juiste zorgtype en specialisme. In die duplicaten wordt het ziekenhuisnummer aangepast zodat altijd te herkennen is dat het om een duplicaat gaat. Tevens wordt in het duplicaatrecord de postcode aangepast zodanig dat het totaal aantal opnamen per postcode voor het hele ziekenhuis kloppend is. In principe worden bij een ziekenhuis waarvoor records ontbreken eerst bestaande records uit het eigen ziekenhuis gedupliceerd, wanneer dit niet mogelijk is wordt gebruik gemaakt van records van vergelijkbare ziekenhuizen (bijv. voor een academisch ziekenhuis worden records gedupliceerd uit de andere academische ziekenhuizen).

Deze gegenereerde records gebruikt het CBS niet bij de koppeling aan de GBA, omdat van deze records geen werkelijke persoonsgegevens bekend zijn. Om deze opnamen wel te laten meetellen bij het afleiden van statistieken op opnameniveau wordt een extra ophoogfactor aan de werkelijke records toegevoegd, de ‘*Ophoogfactor missing records*’. Deze ophoogfactor corrigeert dus voor de ontbrekende records in de LMR. Bij het berekenen van deze ophoogfactor wordt uitgegaan van de wijze waarop Prismant de duplicaatrecords genereert. In plaats van het dupliceren van een record uit de groep bijpassende werkelijke records wordt door het CBS aan elk werkelijk record uit die groep een klein gewicht meegegeven. De gewichten worden op deze manier verspreid over meerdere records, zodat het

niet kan voorkomen dat door toeval een record van een patiënt met een zeldzame ziekte een hoog gewicht krijgt. In tegenstelling tot Prismant wordt bij deze methode echter niet gecorrigeerd voor de postcode. In de finale herweging wordt wel een regionale correctie meegenomen.

4.1 Voorbereiding

Om de verwerkingsnelheid te verhogen worden de benodigde LMR-bestanden eerst ingekort door alle variabelen die voor het bepalen van de ophoogfactoren niet nodig zijn te verwijderen (stap 1)¹. Het gaat hier om de zogenaamde oposbestanden van de LMR waarin de administratieve opnamegegevens zijn opgenomen. Deze zijn na de koppeling onderverdeeld in aparte oposbestanden voor de uniek gekoppelde, meervoudig gekoppelde en niet gekoppelde LMR-records, alsmede het oposbestand met gegenereerde records. Na het inkorten worden de oposbestanden met uniek, meervoudig en niet gekoppelde records samengevoegd tot een totaalbestand *Opos* waarin alle LMR-records staan met uitzondering van de gegenereerde records. Het oposbestand met de gegenereerde records blijft een apart bestand: *Opos-gegenereerd*. De procedure *zetoposklaar* zorgt ervoor dat deze bestanden worden klaargezet.

4.2 Bepaling weeggroepen

Om de *Ophoogfactor missing records* te kunnen bepalen worden de oorspronkelijk geregistreerde records en de gegenereerde records zodanig in weeggroepen verdeeld dat de records van ziekenhuizen waarvoor door Prismant records gegenereerd zijn en de records van ziekenhuizen waarvan de gegenereerde records afkomstig zijn in dezelfde weeggroep worden ondergebracht. Voor elk van de weeggroepen wordt het totale aantal records en het aantal oorspronkelijk geregistreerde records bepaald. De verhouding van beide aantallen levert de *Ophoogfactor missing records* op, welke aan de oorspronkelijk geregistreerde records wordt toegekend. Records van ziekenhuizen die niet bij het generatieproces betrokken zijn krijgen de waarde 1,00 als *Ophoogfactor missing records* toegekend. De weeggroepen worden zo samengesteld dat records van eenzelfde ziekenhuis aan eenzelfde groep worden toegewezen. Als records van ziekenhuis A gebruikt worden voor het genereren van records voor ziekenhuis B en ziekenhuis C, dan zal de bijbehorende weeggroep de records van de ziekenhuizen A, B en C bevatten.

De weeggroepen worden per registratiejaar, zorgtype en specialisme gemaakt.

¹ De stappen die nodig zijn voor het berekenen van de ophoogfactor voor ontbrekende records zijn ingebouwd in een programma dat los staat van dat van de koppelkansen, de nummering van de stappen begint daarom hier weer bij 1. De stappen voor de finale-weging (zie hoofdstuk 5) zijn in hetzelfde programma als de factor voor ontbrekende records ingebouwd, daar loopt de nummering daarom wel door.

De bepaling van de groepen loopt via de procedure *bepaalgroepen* in vier stappen, Eerst vindt er een sortering plaats van de bestanden *Opos* en *Opos-gegenereerd* op basis van de sorteersleutel: registerjaar, zorgtype, opnamedatum, opname-uur, ontslaguur, geslacht, geboortjaar, geboortemaand, verpleegduur en specialisme. De sleutel van het oorspronkelijk geregistreerde record en de sleutel van het gegenereerde record komen overeen en kan daarom gebruikt worden bij het bepalen van de oorsprong van de gegenereerde records.

Stap 2 verwijdert het geringe aantal records in het bestand *Opos* met een gelijke sorteersleutel, aangezien de sleutels van deze records niet eenduidig naar een record in het bestand *Opos-gegenereerd* verwijzen. Dit levert het bestand *Enkelen* op. In de praktijk blijkt het verwijderen van de dubbele records het proces van de groepsvorming niet te verstoren.

Stap 3 gaat op basis van de gehanteerde sorteersleutel na welk bestaand record met welk gegenereerd record koppelt en schrijft bij een geslaagde koppeling de sleutel en de instellingsnummers van het bestaande record en het gegenereerde record weg in een apart bestand. Elke geslaagde koppeling verbindt een donorinstelling met een ontvangende instelling. Dit levert een directe verbinding op tussen instellingen. Als het bestaande record en het gegenereerde record uit dezelfde instelling afkomstig zijn, dan levert dit een directe koppeling op van de instelling met zichzelf.

De in stap 3 weggeschreven records vormen de input voor stap 4 bij het bepalen van de weeggroepen. Per registratiejaar, zorgtype en specialisme worden er disjuncte groepen van instellingen gevormd. Instellingen die volgens stap 3 direct verbonden zijn worden in eenzelfde groep geplaatst. Daaraan worden toegevoegd alle instellingen die via een keten van direct verbonden instellingen verbonden kunnen worden. Een instelling die alleen met zichzelf direct verbonden is, vormt een groep op zich. Elke groep krijgt een groepsnummer.

Stap 5 maakt een lijst waarin per registratiejaar, zorgtype en specialisme de verschillende instellingen die een rol spelen bij het genereren van records zijn voorzien van een groepsnummer.

4.3 Bepaling van de ophoogfactor missing records

De gewichten van de *Ophoogfactor missing records* zijn nu eenvoudig te berekenen door per groep de verhouding te nemen tussen het totale aantal records en het aantal niet-gegenereerde records. Records van ziekenhuizen die niet bij het genereren voor een bepaald zorgtype en specialisme betrokken zijn krijgen groepsnummer 0 en gewicht 1,00.

De bepaling van de missing records gewichten begint met het sorteren op registratiejaar, zorgtype, specialisme en instellingsnummer van *Opos-gegenereerd*, het opnamebestand van de gegenereerde records en *Opos*, het opnamebestand van de bestaande records.

In stap 6 worden op basis van de twee gesorteerde bestanden tellingen gemaakt van de aantallen records per registratiejaar, zorgtype, specialisme en instellingsnummer.

Stap 7 voegt hier de groepsnummers aan toe, hetgeen tot een tussenbestand leidt, en telt de aantallen records per registratiejaar, type zorg, specialisme en groep. Dit gebeurt voor de bestaande records en voor de gegenereerde records.

In stap 8 worden ten slotte de *missing records gewichten* bepaald, per registratiejaar, type zorg, specialisme en instelling op basis van de uitgevoerde groepstellingen.

5. De finale herweging

In de voorgaande hoofdstukken is uiteengezet op welke wijze de koppelkansgewichten zijn bepaald die corrigeren voor de onvolledige koppeling van LMR-records aan de GBA vanwege meervoudige koppeling en onvolledige koppelsleutels en hoe de missing records gewichten zijn bepaald die corrigeren voor de onvolledigheid van de registratie van opnamen. Toepassing van het product van beide gewichten op de uniek gekoppelde records zou in het ideale geval de totale populatie van ziekenhuisopnamen moeten representeren. Dit is echter nog niet geheel het geval, dit kan onder andere komen door het niet koppelen van records als gevolg van administratieve fouten in de koppelvariabelen. (zie ook Projectgroep ontwikkeling GSB, 2003, hoofdstuk 5). Daarom wordt nog een derde weging uitgevoerd om voor deze gemiste koppelingen en voor toevalsfluctuaties bij het koppelproces te corrigeren. Hierbij wordt ook de regionale verdeling goedgemaakt, waar bij de *missing records* factor geen rekening mee gehouden is.

Bij deze *Finale herweging* worden de uniek gekoppelde records eerst gewogen met de *Ophoogfactor koppelkans* en de *Ophoogfactor missing records*. Vervolgens worden zij herwogen naar de totale populatie van LMR-records op basis van de variabelen geslacht, leeftijd, specialisme, verpleegduur, hoofddiagnose en regio. In bijlage 2 staan de gebruikte (klasse-)indelingen van deze variabelen vermeld. Voor deze herweging wordt de methode van Iterative Proportional Fitting gebruikt.

5.1 Voorbereiding

Als voorbereiding op de laatste fase van de ophoging wordt de procedure *startfinaleweging* uitgevoerd.

Stap 9 maakt een ingedikt populatiebestand op basis van alle oposbestanden per jaar, inclusief de gegenereerden. Dit bestand dient als referentie waarnaar de uniek gekoppelde records worden opgehoogd. Voor de indikking van dit bestand is de variabele *jglsvhzr* afgeleid. De letters verwijzen achtereenvolgens naar de variabelen jaar van registratie, geslacht, leeftijdsklasse, specialismegroep, verpleegduurklasse, hoofddiagnose, zorgtype, en regio. Per combinatie van deze variabelen wordt het totale aantal records vastgelegd. Deze totalen vormen de randtotalen waarnaar de uniek gekoppelde records worden opgehoogd.

In stap 10 wordt van de uniek gekoppelde records het totale voorlopige ophooggewicht bepaald per combinatie van de acht bij stap 9 genoemde variabelen, op basis van de individuele ophooggewichten. Deze individuele ophooggewichten

worden berekend door vermenigvuldiging van de *Ophoogfactor koppelkans* en de *Ophoogfactor missing records*. Dit individuele ophooggewicht wordt in het vervolg gebruikt als startgewicht bij de verdere ophoging (finale herweging).

Op basis van de variabele *jglsvhzzr* kunnen een aantal stratificaties (indelingen) worden afgeleid. De fijnste stratificatie bestaat uit alle categorieën van de variabele *jglsvhzzr*. Deze indeling is echter niet te gebruiken wegens onvoldoende celvulling. In plaats daarvan wordt daarom gebruikt gemaakt van een aantal verschillende stratificaties op basis van steeds andere combinaties van de variabelen waaruit *jglsvhzzr* is opgebouwd. Technisch wordt een combinatie aangegeven door een reeks van enen en nullen waarbij de een aangeeft dat een variabele tot de combinatie behoort en een nul dat de variabele niet tot de combinatie hoort. Elke mogelijke combinatie kan dienen als een stratificatie voor de ophoging. Voor de uiteindelijke ophoging wordt hier gebruik gemaakt van een viertal verschillende stratificaties, waarbij steeds twee of drie van de bij de ophoging betrokken variabelen voorkomen. De variabelen registratiejaar en zorgtype, die in eerste instantie wel in de bestanden zijn meegenomen, zijn uiteindelijk niet gebruikt voor de ophoging. Registratiejaar is niet nodig omdat de ophoging per registratiejaar plaatsvindt en deze variabele dus voor alle records gelijk is. Zorgtype is buiten beschouwing gelaten, omdat het onderscheid tussen dagverpleging en klinische opnamen al is af te leiden uit de verpleegduur, die wel wordt meegenomen in de ophoging. Van de overige variabelen zijn de volgende stratificaties (combinaties van variabelen) bij de ophoging gebruikt:

geslacht × leeftijdsklasse × verpleegduurklasse

geslacht × leeftijdsklasse × hoofddiagnosegroep

specialisme × verpleegduurklasse

hoofddiagnosegroep × regio

Bij de keuze van deze stratificaties is gelet op de celvulling en op wensen met betrekking tot het publicatieschema

In stap 11 worden de populatieaantallen voor elk van deze stratificaties bepaald door aggregatie van de met behulp van stap 9 berekende aantallen. Stap 11 levert aldus de benodigde randtotalen voor de ophoging.

5.2 Weging

In het eenvoudige geval van ophoging op basis van één vooraf of achteraf gekozen stratificatie krijgen alle elementen (dat zijn in dit geval de uniek gekoppelde records) in een stratum een ophoogfactor die gelijk is aan de verhouding van de populatieomvang en de omvang van de op te hogen elementen in het stratum. De omvang van de op te hogen elementen kan daarbij het aantal elementen zijn, maar ook, zoals in dit geval, de som van de eerder bepaalde startgewichten van deze elementen. In het laatste geval levert deze procedure een correctiefactor op en is het uiteindelijke gewicht gelijk aan het startgewicht maal de nieuw bepaalde

correctiefactor. Per stratum is de som van de eindgewichten gelijk aan de populatieomvang.

In het geval er meerdere stratificaties worden toegepast zijn er een aantal verschillende weegmethodes, die in de praktijk min of meer dezelfde resultaten opleveren. Gekozen is de methode van de Iteratieve Proportionele Fitting (IPF), ook bekend onder de namen 'Raking Ratio estimation' en 'Rassen' (Deville et al, 1993). Een voordeel van deze methode is dat de uiteindelijke gewichten positief zijn. Bij deze methode vindt er eerst een weging plaats van de uniek gekoppelde records op basis van de eerste stratificatie volgens de in vorige alinea beschreven ophoogwijze. Dit levert een (nieuw) gewicht op dat als startgewicht dient voor de volgende stap. Daarin vindt op dezelfde wijze een ophoging plaats naar de populatietotalen van de tweede stratificatie, waarbij opnieuw de gewichten worden aangepast. Nadat alle stratificaties aan de beurt zijn geweest vindt er een nieuwe ronde plaats: de eindgewichten van de eerste ronde dienen dan weer als startgewicht voor de eerste stratificatie, worden aangepast en gaan naar de tweede stratificatie etc., etc. Dit proces convergeert in het algemeen na een aantal weegronden (iteraties) en levert dan gewichten op die voor elk van de stratificaties leiden tot de gewenste populatieaantallen. Als dit niet het geval is dan moeten de strata nog wat worden ingedikt.

De berekening van de gewichten wordt uitgevoerd in stap 12 en stap 13. Op basis van de gekozen stratificaties worden deze stappen automatisch gegenereerd. Nadat er voldoende iteratieslagen zijn uitgevoerd voegt stap 14 de gewichten aan de records toe, althans aan een sleutelbestand dat één op één correspondeert met het bestand met uniek gekoppelde records. Vanuit dit sleutelbestand worden de gewichten tot slot toegevoegd aan de uniek gekoppelde oposrecords.

De op deze wijze bepaalde gewichten zijn de eindgewichten van de ophoogfactor voor LMR-GBA-statistieken op opnameniveau.

5.3 Vaststelling technische kwaliteit weging

Om te bepalen of het aantal iteraties voldoende is geweest, worden met behulp van stap 12 de geschatte populatieaantallen (de gewogen aantallen opnamen) voor de verschillende strata bepaald. Deze worden vervolgens vergeleken met de overeenkomstige populatieaantallen (het totaal aantal opnamen in de oorspronkelijke LMR). Deze vergelijking gebeurt handmatig. Om dit te vergemakkelijken worden de geschatte en de eerder vastgestelde populatieaantallen per stratum met behulp van stap 15 naast elkaar gezet.

Bij de hier uitgevoerde weging bleken de gewichten na 32 iteratieslagen voldoende te convergeren: de relatieve afwijkingen van de geschatte aantallen en de populatieaantallen bedroegen toen minder dan 0,001%.

6. Literatuur

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.

Van Sijl, Marije & De Ree, Jos (2005). Ophogen op persoonsniveau van gegevens van de Landelijke Medische Registratie gekoppeld met de GBA. Interne CBS-nota, nr. 0160-05-SOO, 17 mei 2005. CBS, Voorburg.

Bijlage 1.

Indeling in geboortelandgroepen

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	

Bijlage 2.

Beschrijving van de variabelen die bij de finale herweging zijn gebruikt

Leeftijd:

De leeftijd betreft de leeftijd aan het begin van het registratiejaar en is ingedeeld in de volgende leeftijdsklassen:

- 0 jaar
- 1-14 jaar
- 15-29 jaar
- 30-44 jaar
- 45-59 jaar
- 60-64 jaar
- 65-69 jaar
- 70-74 jaar
- 75-79 jaar
- 80-89 jaar
- 90 jaar en ouder

De leeftijdsklassen zijn bepaald op basis van het geboortjaar en registratiejaar. Tot de nuljarigen zijn hier gerekend degenen die in het registratiejaar en in het jaar daaraan voorafgaand zijn geboren. Deze keuze leidt tot een betere celvulling in de laagste leeftijdsklasse. De leeftijdsklassen zijn hier, in verband met de benodigde celvulling voor de finale herweging, in het algemeen wat grover dan de leeftijdsklassen gehanteerd bij het bepalen van de gewichten voor de koppelkans. Dit is echter geen bezwaar voor analyses met gedetailleerde leeftijdsklassen, omdat de finale herwegingsgewichten relatief klein zijn ten opzichte van de koppelkansgewichten.

Geslacht:

Man/vrouw

Verpleegduur:

De volgende klassen zijn gebruikt voor de verpleegduur:

- 1 dag
- 2-4 dagen
- 5-7 dagen
- 8-14 dagen
- 15 dagen en langer

Specialismegroep:

De in de ziekenhuizen uitgeoefende specialismen, waarbij de kleinere specialismen zijn samengevoegd in groepen op basis van vergelijkbaarheid. De volgende specialismegroepen zijn onderscheiden:

- Inwendige geneeskunde, alg.
- Cardiologie
- Longziekten
- Reumatologie
- Gastro-enterologie
- Anesthesiologie
- Heelkunde, algemeen
- Urologie
- Verloskunde en gynaecologie
- Kindergeneeskunde, algemeen
- Psychiatrie
- Neurologie
- Dermatologie, algemeen
- Keel-, neus-, oorheelkunde
- Oogheelkunde
- Klinische geriatrie
- Mondziekten/kaakchirurgie, algemeen
- Neonatologie
- Allergologie, Endocrinologie, Hematologie, Hepatologie, Immunologie, Nefrologie, Intensivisme, inwendige geneeskunde, Pijnbestrijding
- Orthopedie, Neurochirurgie, Plastische chirurgie, Cardio-pulmonale chirurgie, Vaatchirurgie, Kinderchirurgie, Thorax chirurgie, Intensivisme, niet inwendige geneeskunde, Traumatologie
- Radiodiagnostiek, Radiotherapie
- Zenuw- en zielsziekten, Revalidatie, Radiologie, Klinische genetica
- Restcategorie, bevat alle niet genoemde specialismen die wel op de codelijst van Prismant voorkomen

Hoofddiagnosegroep:

Indeling van de hoofddiagnose in de volgende 18 groepen (ICD-9 hoofdgroepen):

- 1 Infectieuze en parasitaire ziekten
- 2 Nieuwvormingen
- 3 Ziekten bloed, bloedbereid.org., immuniteitsstoornissen
- 4 Endocriene, voedings-, stofwisselingsziekten\
- 5 Psychische stoornissen
- 6 Ziekten van zenuwstelsel en zintuigen
- 7 Ziekten van hart en vaatstelsel
- 8 Ziekten van de ademhalingsorganen
- 9 Ziekten van de spijsverteringsstelsel
- 10 Ziekten van huid en onderhuids bindweefsel

- 11 Ziekten van spieren, beenderen en bindweefsel
- 12 Ziekten urinewegen en geslachtsorganen
- 13 Complicaties van zwangerschap, bevalling, kraambed
- 14 Aandoeningen van de perinatale periode
- 15 Aangeboren afwijkingen
- 16 Symptomen, onvolledig omschreven ziektebeelden
- 17 Ongevalsletsels en vergiftigingen
- 18 Factoren die de gezondheidstoestand beïnvloeden en contacten met gezondheidsdiensten

Regio:

Als regioindeling is gebruikt de indeling in 40 COROP-gebieden.

Bijlage 3.

Stappenoverzicht van de ophoogprogramma's

Berekening Ophoogfactor koppelkans:

Stap	Inhoud
1	Maakt GBA-basisbestand met alleen de voor de ophoging relevante variabelen, voegt records samen waarin een andere mutatie dan op de koppelvariabelen heeft plaatsgevonden en voegt hieraan het dagnummer en de geboortelandgroep toe.
2	Maakt een bestand met het aantal personen per postcode, geslacht, geboortjaar en geboortelandgroep, en een bestand met per geboortelandgroep, geslacht en geboortjaar het totale aantal personen met een geldige geboortedag, het aantal personen met een ongeldige geboortedag en de aantallen personen per geldige geboortedag.
3	Berekent de deelfactor ' <i>koppelkans</i> ' per postcode, geboortjaar en geboortelandgroep.
4	Bepaalt welk deel van de personen per geslacht, geboortjaar en geboortelandgroep een geldige geboortedatum heeft.
5	Bepaalt de stand van de bevolking naar geslacht en geboortjaar op het eind van het registratiejaar.
6	Bepaalt welke personen uit het resultaatbestand van stap 5 uniek koppelbaar zijn.
7	De deelfactoren ' <i>koppelkans</i> ' en ' <i>kansloos</i> ' worden toegevoegd.
8	De gewichten op basis van ' <i>koppelkans</i> ' en ' <i>kansloos</i> ' worden berekend.
9	De factor ' <i>tweeling</i> ' wordt berekend op basis van de verhouding tussen de werkelijke aantallen personen en de met ' <i>koppelkans</i> ' en ' <i>kansloos</i> ' opgehoogde aantallen.
10	Splitst LMR-oposbestand in jaarbestanden indien met samengevoegde jaren wordt gewerkt.
11	Rekent de <i>Ophoogfactor koppelkans</i> uit en koppelt die aan de uniek gekoppelde LMR-oposrecords.
12	Rekent de koppelkansen per postcode, geslacht, geboortjaar en geboortelandgroep uit voor het begin en het einde van elk jaar. (deze stap wordt niet gebruikt voor de hier beschreven ophoging op opnameniveau, maar voor de ophoging op persoonsniveau)

Berekening Ophoogfactor missing records en finale herweging:

Stap	Inhoud
1	Kort de LMR-oposbestanden in door de niet gebruikte variabelen weg te laten en verdeelt de bestanden in een bestand met uniek, meervoudig en niet gekoppelde records en een bestand met gegenereerde records.
2	Verwijdert records die op de gebruikte sorteervariabelen (registerjaar, zorgtype, opnamedatum, opname-uur, ontslaguur, geslacht, geboortejaar, geboortemaand, verpleegduur en specialisme) volledig overeenkomen.
3	Bepaalt op welke bestaande records de gegenereerde records zijn gebaseerd.
4	Bepaalt de groepen waarin gewogen wordt om de <i>Ophoogfactor missing records</i> te bepalen.
5	Maakt bestanden met het nummer van de te gebruiken weeggroepen erin.
6	Telt het aantal werkelijke en gegenereerde records per jaar, zorgtype, specialisme en instelling.
7	Voegt de nummers van de weeggroepen toe aan de uitvoer van stap 6 en telt de aantallen per registratiejaar, zorgtype, specialisme en weeggroep.
8	Bepaalt de <i>Ophoogfactor missing records</i> .
9	Telt voor de finale herweging het totaal aantal LMR-oposrecords per registratiejaar, geslacht, leeftijdsklasse, specialismegroep, verpleegduurklasse, hoofddiagnosegroep, zorgtype en regio.
10	Bepaalt de voorlopige (start-)gewichten voor de uniek gekoppelde records door de <i>Ophoogfactor koppelkans</i> en de <i>Ophoogfactor missing records</i> te vermenigvuldigen.
11	Bepaalt de populatieaantallen per gekozen stratificatie.
12	Berekent per stratum de in de vorige stratificatie berekende gewichten.
13	Berekent de nieuwe gewichten per stratum.
14	Voegt de uiteindelijke gewichten toe aan een sleutelbestand. Vanuit dit sleutelbestand worden de eindgewichten later toegevoegd aan het uniek gekoppelde LMR-oposbestand.
15	Maakt een controlebestand, waarin de opgehoogde en werkelijke aantallen per stratum naast elkaar worden gezet.