

Koppeling van LMR- en GBA-gegevens: methode, resultaten en kwaliteitsonderzoek

***Projectgroep ontwikkeling
GezondheidsStatistisch Bestand***

Auteurs:

***A. de Bruin
E.I. de Bruin
A. Gast
J.W.P.F. Kardaun
M. van Sijl
G.C.G. Verweij***



Verklaring der tekens

.	= gegevens ontbreken
*	= voorlopig cijfer
x	= geheim
–	= nihil
–	= (indien voorkomend tussen twee getallen) tot en met
0 (0,0)	= het getal is minder dan de helft van de gekozen eenheid
niets (blank)	= een cijfer kan op logische gronden niet voorkomen
2002–2003	= 2002 tot en met 2003
2002/2003	= het gemiddelde over de jaren 2002 tot en met 2003
2002/'03	= oogstjaar, boekjaar, schooljaar enz., beginnend in 2002 en eindigend in 2003
1992/'93–2002/'03	= boekjaar enz. 1992/'93 tot en met 2002/'03

In geval van afronding kan het voorkomen dat de totalen niet geheel overeenstemmen met de som der opgetelde getallen.

Verbeterde cijfers in staten en tabellen zijn niet als zodanig gekenmerkt.

Colofon

Uitgever

Centraal Bureau voor de Statistiek
Prinses Beatrixlaan 428
2273 XZ Voorburg

Druk

Centraal Bureau voor de Statistiek
Facilitair bedrijf

Omslagontwerp

WAT ontwerpers, Utrecht

Inlichtingen

Tel.: 0900 0227 (€ 0,50 per minuut)
Fax: (045) 570 62 68
E-mail: infoservice@cbs.nl

Bestellingen

E-mail: verkoop@cbs.nl

Internet

www.cbs.nl

© Centraal Bureau voor de Statistiek,
Voorburg/Heerlen, 2003.
Bronvermelding is verplicht.
Vereenvoudiging voor eigen gebruik of
intern gebruik is toegestaan.

Prijzen zijn excl. administratie- en
verzendkosten.
Prijs: € 20,00
Kengetal: C-160
ISBN 903572818 1
CSB-productnummer: 6008703010



Centraal Bureau voor de Statistiek

Inhoud

Voorwoord	5
Samenvatting	7
1. Inleiding	9
2. Algemene beschrijving van LMR en GBA	10
2.1 LMR	10
2.2 GBA	10
3. Afbakening van de LMR-populatie	11
3.1 Instellingen	11
3.2 Behandelingen	12
3.3 Buitenlanders	12
3.4 Gegeneerde records (missende microgegevens)	12
3.5 Overzicht afbakening LMR-bestanden	14
4. Koppeling van LMR aan GBA-gegevens	15
4.1 Het CKP	15
4.2 Voorbewerking LMR voor koppeling met CKP	15
4.3 Het koppelen van LMR aan CKP	15
4.4 Resultaten na koppelingsstappen I en II	18
4.5 Resultaten na koppelingsstap III	18
4.6 Resultaten van koppeling op alleen postcode-4 sleutel	19
5. Onderzoek naar de kwaliteit van de koppeling	20
5.1 De koppelvariabelen in de LMR	20
5.2 De koppelvariabelen in de GBA	21
5.3 Mogelijke miskoppelingen op basis van de postcode	22
5.4 Vergelijking van overlijdensdata in LMR en in GBA	23
5.5 Onwaarschijnlijke mutaties in de GBA-gegevens tijdens LMR-opnameperiode	25
5.6 Overlappende opnameperioden bij dezelfde persoon	25
5.7 Kenmerken van de niet gekoppelde LMR-records	28
6. Conclusies kwaliteitsonderzoek	32
7. Beschouwing	33
Bijlagen	34
1. Inhoud LMR-bestanden	34
2. Bij koppeling toegevoegde gegevens	36
3. Resultaten overlappende opnamen bij dezelfde instelling	37
Referenties	38

Voorwoord

Bij het vervaardigen van statistieken probeert het Centraal Bureau voor de Statistiek (CBS) in toenemende mate gebruik te maken van gegevens die in de samenleving toch al voor andere doeleinden worden vastgelegd. Met andere woorden: grootschalig statistisch hergebruik van administratieve gegevensbestanden. Op het terrein van de gezondheidszorg bestaat een groot aantal van dergelijke administratieve gegevensbestanden, die echter tot voor kort niet door het CBS konden worden benut voor statistische doeleinden. In 2002 is hierin verandering gekomen toen het CBS startte met de ontsluiting van de Landelijke Medische Registratie (LMR) voor het maken van nationale statistieken over (on-)gezondheid en ziekenhuiszorggebruik. Dit werd mogelijk gemaakt door een samenwerkingsovereenkomst met de Stichting Prismant. De ontsluiting betreft in eerste instantie de koppeling op persoonsniveau van LMR-gegevens aan gegevens uit de Gemeentelijke Basisadministratie (GBA), waarmee in principe ook diverse andere voor de statistiek relevante populatiegegevens gekoppeld kunnen worden. Deze koppeling van LMR- aan GBA-gegevens wordt in dit rapport beschreven.

Door dergelijke koppelingen ontstaan nieuwe, in Nederland nog ongekende, mogelijkheden voor het samenstellen van statistische informatie, zonder dat dit leidt tot additionele administratieve

lastendruk voor de ziekenhuizen. Het uitsplitsen van de resultaten naar sociaal-economische status en het doen van longitudinaal onderzoek zijn daarvan twee belangrijke voorbeelden. Deze rapportage is daarmee een majeure stap vooruit in de richting van een volledig, samenhangend en consistent statistisch beeld van gezondheid en zorggebruik in Nederland.

Op deze plaats past een bijzonder woord van dank aan Prismant. Medewerkers van Prismant zijn intensief betrokken geweest bij de leverantie van de LMR-gegevens aan het CBS, hebben informatie verstrekt over de LMR-registratie en de conceptversie van dit rapport van commentaar voorzien. Het CBS is Prismant erkentelijk voor haar inzet, informatie en adviezen, zonder welke dit onderzoek niet mogelijk was geweest.

De Directeur-Generaal van de Statistiek

Ir. drs. R.B.J.C. van Noort

Voorburg, december 2003

Samenvatting

In dit rapport worden methodologie, resultaten en kwaliteit beschreven van de door het CBS ontwikkelde koppeling van gegevens uit de Landelijke Medische Registratie (LMR) met bevolkingsgegevens uit de Gemeentelijke Basisadministratie (GBA). De Stichting Prismant heeft hiervoor gegevens uit de LMR ter beschikking gesteld aan het CBS. De LMR bevat gegevens over uit het ziekenhuis ontslagen patiënten. Behalve opname- en ontslaggegevens worden ook gegevens over diagnoses en medische verrichtingen vastgelegd. De GBA bevat gegevens over alle in Nederland in het bevolkingsregister ingeschreven personen. Het CBS krijgt gegevens uit de GBA voor het maken van statistische informatie, die onder andere tot stand komt door koppeling met andere gegevensbronnen. De koppeling op persoonsniveau van LMR- met GBA-gegevens wordt beschouwd als een belangrijke stap op weg naar de ontwikkeling van een Gezondheids-Statistisch Bestand (GSB). In het GSB worden gegevens over gezondheid en zorggebruik uit verschillende bronnen geïntegreerd met diverse populatiegegevens op sociaal-economisch terrein. De GBA vormt de ruggengraat voor de hiervoor benodigde koppelingen.

Alvorens tot koppeling van LMR- en GBA-gegevens over te gaan is de populatie van instellingen en opnamen in de LMR als volgt uniform afgebakend:

- Ten aanzien van de instellingen zijn de gegevens afgebakend tot alle academische en algemene ziekenhuizen en alle categorale ziekenhuizen met uitzondering van revalidatie- en astmacentra (long-stay ziekenhuizen) en epilepsieklinieken (nemen niet deel aan LMR).
- Op niveau van de opnamen zijn uitgesloten de opnamen met behandelingen die volgens de regels van de LMR niet geregistreerd hadden moeten worden (poliklinische bevalling, psychiatrische deeltijdbehandeling, revalidatie-dagbehandeling). Opnamen van buitenlandse patiënten zijn ook niet meegenomen omdat bij de koppeling uitgegaan wordt van de in Nederland in de GBA ingeschreven populatie. Tenslotte zijn de gegenereerde records, die Prismant aan de LMR-bestanden heeft toegevoegd ter correctie voor de ca. 1% niet geregistreerde (missende) opnamen in de LMR, vóór de koppeling uit de bestanden verwijderd omdat dit geen records zijn van 'echte' opnamen.

De LMR-gegevens van de jaren 1995–2001 zijn vervolgens in een aantal stappen gekoppeld met gegevens uit de GBA. Hierbij is uitgegaan van de primaire koppelvariabelen opnamedatum, geboortedatum, geslacht en postcode (postcode eerst op 6 posities, wanneer beschikbaar, in koppelingstap I; vervolgens de nog niet gekoppelde records op postcode 4 posities in koppelingstap II) en daarna is ook de overlijdensdatum van in het ziekenhuis overleden LMR-patiënten als koppelvariabele meegenomen (koppelingstap III). Bij de koppeling is uitgegaan van de GBA-records die geldig waren op de opnamedatum van het desbetreffende LMR-record, waarbij een marge van ± 30 dagen is toegelaten.

Het eindresultaat na deze drie koppelingstappen is dat 87,6% van de LMR-records uniek gekoppeld kon worden aan de GBA, bij 8,7% was sprake van een meervoudige koppeling en 3,6% van de records kon niet gekoppeld worden.

Als alleen gekoppeld wordt op viercijferige postcode en geboortedatum, geslacht en opnamedatum leidt dit tot 84,1% uniek gekoppelde, 12,2% meervoudig gekoppelde en eveneens 3,6% niet gekoppelde LMR-records.

De gevonden percentages zijn plausibel. Uitgaande van een administratieve foutenruis van ca. 2% in de GBA (geschat op basis van onderzoek) en een even zo groot percentage in de LMR

(aanname) dan is een percentage van 3,6% niet-gekoppelde records zeker niet hoger dan wat verwacht mag worden. De meervoudig gekoppelde records zijn een natuurlijk gevolg van het beperkte discriminerende vermogen van de koppelsleutel, met name bij de records met alleen een viercijferige postcode (betreft driekwart van alle LMR-records).

De niet gekoppelde records omvatten ook de 'gemiste koppelingen', de onterecht niet gekoppelde records. Gezien het lage percentage niet gekoppelde records zal het effect op de statistische uitkomsten in het algemeen klein zijn, met name wanneer deze gebaseerd worden op de totale groep van uniek-gekoppelde records. De niet gekoppelde LMR-records betreffen wel een enigszins selectieve groep. Vergeleken met de overige LMR-records betreffen de niet gekoppelde records verhoudingsgewijs vaker 0-jarigen, jongvolwassenen en zeer ouderen; mannen; klinische opnamen, psychische stoornissen en ongevalsletsels als hoofd-diagnose en zijn de patiënten vaker afkomstig uit de grote steden in het westen des lands. De verschillen zijn weliswaar duidelijk waarneembaar maar aan de andere kant ook niet zeer groot en niet sterk fluctuerend in de tijd. Hoewel men zich bij de analyse van de gekoppelde gegevens bewust dient te zijn van deze verschillen, is het – gezien de relatief beperkte mate van selectiviteit en omvang van de niet-gekoppelde groep – de verwachting dat de invloed hiervan op de meeste analyses klein zal zijn.

Verder is onderzoek gedaan naar de mogelijke omvang van de 'miskoppelingen', de onterecht gekoppelde records. Dit is onder andere bestudeerd door de resultaten van de koppeling op postcode-6 niveau te vergelijken met die op postcode-4 niveau van de LMR-records met een potentieel geldige postcode-6. Hieruit is naar voren gekomen dat minder dan 3% van de uniek op postcode-4 niveau gekoppelde records miskoppelingen kunnen betreffen. Het gaat hier om een bovengrens, omdat deze 3% ook nog een onbekend deel juiste koppelingen bevat. Het aandeel miskoppelingen is verder onderzocht door bij de gekoppelde records de omvang van inconsistenties in het overlijdensgegeven en inconsistenties in de vorm van overlappende opnamen na te gaan. Bij het onderzoek naar de overlijdensgegevens is de overlijdensdatum in de LMR van in het ziekenhuis overleden patiënten vergeleken met die zoals geregistreerd in de GBA. Gebleken is dat bij 1,5% van de in de LMR overleden patiënten er een substantieel verschil in overlijdensdatum is bij de aangekoppelde GBA-records; dit hoeven overigens ook niet altijd miskoppelingen te betreffen. Verder is vastgesteld dat er bij minder dan 0,1% van de uniek gekoppelde records sprake is van overlappende opnamen bij dezelfde persoon, wijzend in de richting van miskoppeling. Als alle onderzoeken naar miskoppelingen samen worden genomen, dan kan geconcludeerd worden dat het miskoppelingspercentage waarschijnlijk in de orde van grootte van 1% ligt en in ieder geval minder dan 3% bedraagt. Op grond hiervan kan gesteld worden dat de miskoppelingen van beperkte omvang zijn.

Samengevat kan geconcludeerd worden dat, gezien het slechts in beperkte mate vóórkomen van miskoppelingen en gemiste koppelingen, de in dit rapport beschreven koppeling van de LMR met de GBA van goede kwaliteit is en een voldoende basis vormt voor statistische analyses.

Aanvullende koppelingmogelijkheden zijn nog onderwerp van studie. Het gaat hier met name om de mogelijkheid om meervoudige en niet gekoppelde records 'bij te koppelen' bij de al uniek-gekoppelden op basis van overeenkomstige waarden op secundaire koppelingvariabelen, zoals bijvoorbeeld het ziekenhuisspecifieke patiëntnummer en de huisartscode in de LMR. In het algemeen zullen dergelijke methoden pas in de productie

worden toegepast als hiermee het merendeel van de nu niet uniek gekoppelde records kunnen worden gekoppeld. Anders wordt de ophoging van de gegevens zeer complex, omdat de koppelingen

van bijgekoppelde records en de al eerder gekoppelde records op zeer verschillende wijzen tot stand zijn gekomen.

1. Inleiding

De ontwikkeling van een GezondheidsStatistisch Bestand (GSB) is één van de doelstellingen van het Strategisch Project Zorg van het CBS. Met het GSB wordt beoogd om door middel van koppeling van verschillende gegevensbronnen geïntegreerde informatie over gezondheid samen te stellen. Het gaat hierbij om de gezondheidstoestand van de bevolking als geheel, alsook van specifieke doelgroepen daarbinnen, determinanten voor gezondheid en zorggebruik en de vraag naar en het gebruik van zorgvoorzieningen. De verwachting is dat de koppeling van bestaande gegevensbronnen op microniveau een belangrijke meerwaarde oplevert voor de informatievoorziening op gezondheidsterrein doordat voorheen gescheiden informatiestromen nu in combinatie geanalyseerd kunnen worden. Behalve dat verschillende attributen met betrekking tot zorggebruik, gezondheidstoestand, determinanten, en demografische en sociaal-economische achtergrond op persoonsniveau bij elkaar gebracht worden, is het ook mogelijk om met het GSB bevolkingsgroepen in de tijd te volgen en zo longitudinale statistische informatie samen te stellen. De Gemeentelijke Basisadministratie (GBA), waaruit het CBS gegevens krijgt over alle in Nederland in het bevolkingsregister ingeschreven personen, vormt de ruggengraat voor het GSB. Hiermee worden de verschillende externe en bij het CBS beschikbare registraties op persoonsniveau gekoppeld. De ontwikkeling van het GSB sluit hiermee aan bij, en zal op termijn onderdeel worden van, het al eerder door het CBS in ontwikkeling gebrachte Sociaal Statistisch Bestand (SSB). In het SSB worden populatiegegevens op sociaal-economisch terrein met elkaar geïntegreerd, met ook de GBA als basis.¹⁾

Voor de ontwikkeling van het GSB wordt, om redenen van efficiency en vermindering van enquêtedruk, maximaal gebruik gemaakt van bestaande gegevensbronnen. Met betrekking tot gezondheidsgegevens beschikt het CBS over een beperkt aantal gegevensbronnen op persoonsniveau, waaronder de doodsoorzakenregistratie en de persoonsenquêtegegevens uit het Permanent Onderzoek Leefsituatie (POLS); deze bronnen zijn reeds gekoppeld aan de GBA. Om een GSB te kunnen ontwikkelen dat het brede terrein van de gezondheidszorg bestrijkt, is het echter noodzakelijk om ook externe medische registraties in het GSB op te nemen. Als eerste stap is hierbij gekozen voor de ziekenhuisregistraties, omdat de ziekenhuiszorg een kernonderdeel is van de gezondheidszorg en hiervan registraties beschikbaar zijn met

een hoge dekkingsgraad en met gegevens over zowel zorggebruik als – voor wat betreft de ziekenhuisopnamen – gezondheidstoestand (diagnosen). Bovendien bieden deze gegevens in principe mogelijkheden om bepaalde onderdelen van de primaire waarneming in POLS te vervangen (efficiency).

De ziekenhuisregistraties betreffen de Landelijke Medische Registratie (LMR), met gegevens over ziekenhuisopnamen, en de Landelijke Ambulante Zorg Registratie (LAZR), met gegevens over poliklinische consulten. Beide registraties worden beheerd door de Stichting Prismant. In het kader van een samenwerkingsovereenkomst tussen CBS en Prismant,²⁾ heeft het CBS de beschikking gekregen over microgegevens uit deze registraties.

Dit rapport betreft de koppeling van de LMR met de GBA.³⁾ Het is een methodologische rapportage van de afbakening van de te koppelen populatie, de gevolgde procedures voor de basale koppeling van de gegevens en de onderzoeken die verricht zijn naar de kwaliteit hiervan (hoofdstukken 2–5). In hoofdstuk 6 worden de conclusies van het kwaliteitsonderzoek samengevat en in hoofdstuk 7 wordt tot slot ingegaan op de mogelijkheden voor verdere uitbouw van de koppelingsprocedures en voor ophoging van de gekoppelde gegevens.

Het rapport is een gezamenlijk product van leden van de Projectgroep Ontwikkeling GSB. De koppelingsprocedures zijn ontwikkeld door Fred Gast en uitgevoerd door Edie de Bruin; het kwaliteitsonderzoek van de koppeling is uitgevoerd door Gerard Verweij; Marije van Sijl heeft de nadere afbakening van de populatie beschreven en de eindbestanden afgeleid; de algemene advisering ten aanzien van methodologie en analyse is verzorgd door Jan Kardaun en de projectleiding en samenstelling van de eindrapportage is uitgevoerd door Agnes de Bruin. De vervolgstappen van dit koppelingsproject zijn de ophoging en analyse van de gekoppelde LMR- en GBA-gegevens en het afleiden van statistische uitkomsten. Over deze vervolgstappen zal afzonderlijk worden gerapporteerd en gepubliceerd.

³⁾ Wanneer in dit rapport gesproken wordt over koppeling met 'de GBA', wordt bedoeld de koppeling met aan de GBA ontleende gegevens. Het CBS beschikt niet over de gehele GBA, maar slechts over selecties daaruit.

2. Algemene beschrijving van LMR en GBA

2.1 LMR

De LMR bevat gegevens van ziekenhuisopnamen. Alle academische en algemene ziekenhuizen registreren in de LMR, evenals een aantal categorale ziekenhuizen. In de LMR worden gegevens van zowel dagopnamen als klinische opnamen geregistreerd. Van de opgenomen patiënten worden – naast administratieve gegevens van de patiënt – ook opname- en ontslaggegevens vastgelegd en medische gegevens met betrekking tot diagnoses en verrichtingen. Een deel van de gegevens is facultatief, dat wil zeggen dat niet alle ziekenhuizen aan de registratie hiervan deelnemen. Dit betreft de wijkcode (geregistreerd door circa een kwart van de ziekenhuizen), voorzorg- en nazorggegevens (27–43% deelname) en het ziekenhuisspecifieke patiëntnummer (90–97% deelname). De diagnose- en verrichtinggegevens worden vastgelegd naar de verantwoordelijkheidsperiode van de bij een opname betrokken specialisten. De eerste verantwoordelijkheidsperiode begint bij de opname, bij elke overdracht volgt een nieuwe verantwoordelijkheidsperiode. Per verantwoordelijkheidsperiode is er 1 verantwoordelijk specialist, max. 8 medebehandelaars, 1 primaire diagnose, max. 10 nevendiaagnosen en max. 99 verrichtingen. Het verantwoordelijke specialisme bij ontslag bepaalt de hoofddiagnose en hoofdverrichting van de gehele opname, dit is dus één van de primaire diagnoses, resp. primaire verrichtingen van de verantwoordelijkheidsperiodes. Overigens heeft ca. 96% van de patiënten maar één verantwoordelijkheidsperiode.

De administratieve gegevens worden verzameld door de administratie van een ziekenhuis bij elke opname. Bij ontslag worden de medische gegevens ingevuld door of namens de specialist op het ontslagformulier. Deze gegevens worden door de medische administraties van de ziekenhuizen gecodeerd en vastgelegd. De richtlijnen hiervoor zijn vastgelegd in het handboek LMR.³⁾ De gegevens worden vervolgens naar Prismant gezonden, die een groot aantal rubriek- en relatiecontroles hierop uitvoert. De resultaten hiervan worden teruggekoppeld naar de ziekenhuizen, waarna eventueel correctie plaatsvindt. Wanneer een ziekenhuis deelneemt aan de registratie van facultatieve gegevens, wordt gecontroleerd of deze gegevens ook inderdaad worden aangeleverd.

In 2001 heeft Prismant in samenwerking met een groot aantal ziekenhuizen onderzoek gedaan naar de betrouwbaarheid van de LMR door de informatie in de LMR te vergelijken met de gegevens zoals vastgelegd in de patiëntendossiers. Hieruit is naar voren gekomen dat de betrouwbaarheid in het algemeen goed is: de administratieve persoonsgegevens en opname- en ontslaggegevens stemden in 99% van de gevallen overeen, bij de hoofd diagnoses was de overeenstemming 84%, bij de nevendiaagnosen 87%, bij de verrichtingen 92% en bij de ongevalstoedrachtgegevens 91%.⁴⁾ De hoge overeenstemming van de persoonsgegevens (waaronder viercijferige postcode) hoeft overigens nog niet te betekenen dat het geregistreerde adres hetzelfde adres is als dat wat in de GBA geregistreerd is; men zou in sommige gevallen bijvoorbeeld ook het adres van de verzorger of de financiële administrateur aan het ziekenhuis kunnen opgeven. Ten aanzien van de nevendiaagnosen is geen onderzoek gedaan naar de volledigheid, d.w.z. of ook alle in het patiëntendossier vermelde diagnoses in de LMR geregistreerd staan.

De LMR is sinds 1963 operationeel (toen nog Stichting Medische Registratie geheten) en is in 1991 gereviseerd. Op grond van de CBS-Prismant samenwerkingsovereenkomst²⁾ van februari 2002 levert Prismant jaarlijks LMR-microgegevens aan het CBS, vanaf verslagjaar 1995. Een specificatie van de geleverde gegevens is gegeven in Bijlage 1. Elk LMR-verslagjaar bevat gegevens over de opnamen van in dat jaar uit het ziekenhuis *ontslagen* patiënten. Als startjaar is 1995 gekozen omdat van de GBA, waar-

mee de koppeling is beoogd, vanaf dit jaar gegevens beschikbaar zijn. In september 2002 waren de LMR-gegevens van de jaren 1995–2001 aan het CBS geleverd. De in dit rapport beschreven koppeling is uitgevoerd met de LMR-gegevens van deze verslagjaren.

2.2 GBA

De GBA is een geautomatiseerd persoonsregistratiesysteem van de Nederlandse gemeenten, in werking sinds 1 oktober 1994. De GBA bevat informatie over alle in Nederland ingeschreven personen. Het gaat hierbij om demografische gegevens zoals geslacht, geboortedatum, postcode, geboorteland, familierechtelijke betrekkingen en overlijdensdatum. Het CBS heeft ten behoeve van zijn statistische taak de beschikking over structuurgegevens (stand van de bevolking per 1 januari) en stroomgegevens (mutaties door onder meer verhuizing, geboorte en sterfte) op persoonsniveau. Deze gegevens vormen de basis voor de bevolkingsstatistiek van het CBS. Tien procent van de gegevens worden gebruikt als steekproefkader voor de CBS persoonsenquête. Een gedetailleerde beschrijving van de inhoud van de GBA, de gegevens die het CBS uit de GBA verkrijgt en de verwerking hiervan door het CBS, is elders gegeven.⁵⁾

In 2002 is er onderzoek gedaan naar de kwaliteit van het adresgegeven in de GBA.⁶⁾ In het kader van de gemeenteraadsverkiezingen is bij een steekproef van oproepingskaarten bij 21 gemeenten nagegaan, door persoonlijke controle en overhandiging van de oproepkaart aan de kiesgerechtigde, of het in de GBA geregistreerde adres van de kiesgerechtigde juist is. De resultaten van het onderzoek geven aan dat 97,3% van de personen op het juiste adres is geregistreerd, 1,5% onjuist en in 1,2% van de gevallen kon de juistheid niet vastgesteld worden omdat de persoon niet bereikt werd. Dit resultaat komt sterk overeen met de uitkomsten van onderzoeken uit 1994 en 1998, toen het percentage juiste adressen 97,5%, respectievelijk 98,5% bedroeg. Verder kwam uit het onderzoek naar voren dat jongeren (18–30-jarigen) vaker een onjuist adresgegeven in de GBA hebben dan ouderen. Dit heeft waarschijnlijk te maken met het feit dat jongeren, waaronder veel studenten, vaker verhuizen en minder zorgvuldig zijn ten aanzien van het steeds opgeven van het feitelijke woonadres dan ouderen.

Geconcludeerd kan worden dat het percentage personen van wie het woonadres in de GBA niet juist is geregistreerd, in de orde van grootte van 1–2% ligt. Hierbij wordt de aantekening gemaakt dat hier alleen de in de GBA ingeschrevenen zijn onderzocht en niet de personen die niet in de GBA zijn ingeschreven, maar wel in Nederland woonachtig zijn. Bij het gebruik van GBA-gegevens als koppelingskader moet ook rekening gehouden worden met marges in de opgegeven en feitelijke verhuisdata: iemand die verhuisd kan bijvoorbeeld om praktische redenen een eerdere of latere datum opgeven aan de bevolkingsadministratie.

Ten behoeve van de koppeling met andere registraties heeft het CBS de verkregen GBA structuur- en stroomgegevens verwerkt in een historisch longitudinaal bestand (Centraal Koppelbestand Personen) waarin de demografische wijzigingen zijn opgenomen. Wanneer bijvoorbeeld een persoon is verhuisd, dan komen gegevens van deze persoon twee keer in het bestand voor: één keer met het oude adres en één keer met het nieuwe adres, met de bijbehorende aanvangs- en beëindigingsdatum. Hiermee is het mogelijk voor ieder persoon op elk moment in de tijd de demografische kenmerken te bepalen. Het Centraal Koppelbestand Personen wordt alleen historisch opgebouwd, dat wil zeggen tot en met 1 januari van het lopende jaar.

3. Afbakening van de LMR-populatie

In dit hoofdstuk wordt beschreven op welke wijze en op grond van welke redenen de populatie van instellingen en opnamen, zoals geregistreerd in de LMR, is afgebakend door het CBS, alvorens tot koppeling met de GBA over te gaan. Op het niveau van de instellingen gaat het met name om een nadere afbakening van het type instellingen. Op het niveau van de opnamen gaat het enerzijds om het uitsluiten van enkele behandelingen die volgens de regels van de LMR niet geregistreerd moeten worden en anderzijds om het uitsluiten van opnamen van buitenlandse patiënten. Ten slotte wordt ingegaan op de niet in de LMR geregistreerde opnamen (de incomplete data), die dus niet gekoppeld kunnen worden, en de mogelijke invloed hiervan op analyse-uitkomsten.

3.1 Instellingen

Uitgangspunten voor de afbakening van de populatie van instellingen zijn de mate van deelname per categorie ziekenhuis (algemeen, academisch of categoriaal) en naar verdere typering van het soort ziekenhuis (specialisatie, opnamepatroon). Psychiatrische ziekenhuizen zijn een andere categorie ziekenhuizen; deze registreren niet in de LMR en behoren dus ook niet bij de populatie van instellingen.

De populatie van instellingen wordt afgebakend tot alle academische en algemene ziekenhuizen en alle categoriale ziekenhuizen met uitzondering van de revalidatie-, epilepsie- en astmacentra. Deze afbakening van instellingen komt overeen met de zogenaamde 'landelijke pool' zoals die door Prismant voor LMR-publicaties gehanteerd wordt. De rationale van deze selectie van instellingen wordt hieronder nader aangegeven.

In tabel 1 is een overzicht gegeven van de aantallen ziekenhuizen per jaar, uitgesplitst naar categorie, en hoeveel daarvan deelnemen aan de LMR. In tabel 2 is dit voor de categoriale ziekenhuizen verder uitgesplitst naar type.

Uit tabel 1 blijkt dat de academische en algemene ziekenhuizen allemaal deelnemen aan de LMR. Deze worden daarom allen geselecteerd voor de te koppelen populatie.

Bij de categoriale ziekenhuizen is de deelname 60–79%, de deelname loopt enigszins af in de tijd, maar dit wordt sterk vertekend door de lage totale aantallen, die eveneens afnemen in de tijd. Naast de in de tabellen genoemde categoriale instellingen zijn er in Nederland ook nog ca. 15 revalidatie-instellingen, die in andere CBS-publicaties tot de categoriale instellingen worden gerekend. Deze revalidatie-instellingen doen geen van allen mee aan de LMR en vallen dus ook noodzakelijkerwijs buiten de te selecteren populatie. Uit tabel 2 blijkt dat dit ook geldt voor de 2 à 3 'single specialty' epilepsieklinieken: deze doen geen van allen mee en worden derhalve ook buiten beschouwing gelaten.

Tabel 1
Deelname van academische, algemene en categoriale ziekenhuizen aan LMR, 1995–2001 (excl. revalidatiecentra)

	Academisch		Algemeen		Categoraal	
	totaal	in LMR	totaal	in LMR	totaal	in LMR
1995	9	9 (100%)	110	110 (100%)	14	11 (78,6%)
1996	9	9 (100%)	106	106 (100%)	13	10 (78,6%)
1997	8	8 (100%)	105	105 (100%)	13	10 (76,9%)
1998	8	8 (100%)	104	104 (100%)	12	9 (75,0%)
1999	8	8 (100%)	104	104 (100%)	12	8 (66,7%)
2000	8	8 (100%)	101	101 (100%)	11	7 (63,6%)
2001	8	8 (100%)	97	97 (100%)	10	6 (60,0%)

Daarnaast zijn er enkele categoriale instellingen die dusdanig afwijken van algemene en academische ziekenhuizen, dat besloten is om deze ook niet mee te nemen in de populatie. Het gaat hier om categoriale ziekenhuizen die een duidelijk 'longstay'-karakter hebben. Van deze ziekenhuizen (voornamelijk revalidatie- en astmacentra) doet slechts een klein deel mee aan de LMR en zij worden van oudsher ook niet meegenomen in LMR-publica-

Tabel 2
Deelname van categoriale instellingen aan LMR, naar type, 1995–2001 (excl. revalidatiecentra)

	1995		1996		1997		1998		1999		2000		2001	
	totaal	LMR	totaal	LMR	totaal	LMR	totaal	LMR	totaal	LMR	totaal	LMR	totaal	LMR
Specialisme*														
S+lg/ast/r/rev/diab	1	1	1	1	1	1	1	1	1	1	1	0	0**	n.v.t.
S+lg/ast	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or/rev/r	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Astma	2	2	2	1	2	1	2	1	2	1	2	1	2	1
Sanatorium/astma	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Oogheelkunde	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Reuma	1	0	0**	n.v.t.	0	n.v.t.	0	n.v.t.	0	n.v.t.	0	n.v.t.	0	n.v.t.
Ep/ast/ig/car/aller	1	1	1	1	1	1	1	1	1	1	1	1	0**	n.v.t.
Kraamkliniek	1	1	1	1	1	1	1	1	0**	n.v.t.	0	n.v.t.	0	n.v.t.
Kanker	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Kanker/reuma	1	1	1	1	1	1	0**	n.v.t.	0	n.v.t.	0	n.v.t.	0	n.v.t.
Epilepsie	2	0	2	0	2	0	2	0	2	0	2	0	3	0

*s= sanatorium, lg= longziekten, ast=astma, r= reuma, rev= revalidatie, diab= diabetes, or= orthopedie, ep= epilepsie, car= cardiologie, aller= allergologie.

**j Deze ziekenhuizen zijn opgegaan in andere (algemene of academische) ziekenhuizen. Soms zijn de verschillende afdelingen opgegaan in verschillende ziekenhuizen. Eén afdeling (Epilepsie) is als zelfstandige kliniek doorgegaan onder een andere naam.

ties. Op grond hiervan worden de 3 aan de LMR deelnemende astmacentra (aangegeven met 'Astma' en 'Sanatorium/Astma' in tabel 2) buiten beschouwing gelaten. Als uitzondering wordt het ziekenhuis aangeduid met 'S+lg/ast/rev/diab' dat ook een enigszins 'longstay'- karakter heeft, wél meegenomen. De reden hiervoor is dat onderdelen van dit ziekenhuis vanaf 2000/2001 zijn opgenomen in een academisch ziekenhuis. Als het categorale ziekenhuis dan eerder niet in de populatie zou zijn opgenomen, zou dit leiden tot een vertekening van de opnamecijfers in de tijd.

Verder dient vermeld te worden dat van de wel in de populatie opgenomen kankerklinieken (n=2 t/m 1997 en n=1 vanaf 1998) uitsluitend geaggregeerde totalen van opnamen bekend zijn en geen microgegevens (zie 3.4). Deze instellingen vallen daarom de facto ook buiten de te koppelen populatie, omdat er geen microgegevens zijn waarmee gekoppeld kan worden. Bij analyses op opnameniveau kan hier echter voor gecorrigeerd worden, bijvoorbeeld door weging. Voor analyses op persoonsniveau ligt dit gecompliceerder en zullen de opnamen van deze instellingen wellicht geheel buiten beschouwing worden gelaten.

3.2 Behandelingen

Sommige ziekenhuizen rapporteren bepaalde behandelingen, die volgens de LMR-handleiding niet in de LMR geregistreerd moeten worden. Daarnaast zijn er opnamen die niet meetellen in de adherentie bepalingen van de ziekenhuizen, maar die volgens de regels van de LMR wel geregistreerd moeten worden. Deze opnamen worden niet door alle ziekenhuizen (volledig) geregistreerd in de LMR. In deze gevallen kan het wenselijk zijn om de desbetreffende behandelingen of opnamen in het geheel uit te sluiten van de populatie, omdat de onderrapportage te groot is.

Besloten is om alle opnamen mee te nemen, met uitzondering van de opnamen voor een poliklinische bevalling, psychiatrische deeltijdbehandeling of revalidatie-dagbehandeling. De verschillende typen behandelingen/opnamen die in het kader van deze afbakening van de populatie bekeken zijn, worden hieronder nader besproken.

Poliklinische bevallingen

Bevallingen in het ziekenhuis kunnen klinisch of poliklinisch plaatsvinden. Poliklinische bevallingen horen niet in de LMR geregistreerd te worden, maar het blijkt dat sommige ziekenhuizen poliklinische bevallingen ten onrechte in de LMR registreren bij de dagverpleging. Zodra er complicaties optreden bij een bevalling, is er sprake van een klinische bevalling. Daarom lijkt het aannemelijk dat alleen bevallingen die in dagverpleging geregistreerd staan en waarvan uit de diagnose blijkt dat er geen complicaties waren (diagnose: 'bevalling bij volledig normaal geval'), ten onrechte geregistreerde poliklinische bevallingen kunnen zijn. Daarom is er voor gekozen om deze opnamen uit de bestanden te verwijderen.

Psychiatrische deeltijdbehandeling

Psychiatrische deeltijdbehandeling is een behandeling voor psychiatrische patiënten waarvoor patiënten met een bepaalde regelmaat een halve tot een hele dag naar het ziekenhuis komen. Deze vorm van behandeling hoort niet geregistreerd te worden in de LMR. Enkele ziekenhuizen registreren echter wel psychiatrische deeltijdbehandeling in de LMR, onder dagverpleging. Deze opnamen worden daarom verwijderd uit de bestanden (met selectie: zorgtype = dagverpleging en ontslagspecialisme = psychiatrie). Er wordt hier geselecteerd op ontslagspecialisme en niet op psychiatrische diagnoses. Psychiatrische diagnoses kunnen

namelijk wel rechtmatig voorkomen in de dagverpleging, een behandeling door een psychiater niet.

Revalidatie

Revalidatie kan voorkomen tijdens een klinische opname of als revalidatie-behandeluren. Revalidatie in dagverpleging bestaat volgens de administratieve regelgeving niet. Een heel enkele keer wordt door ziekenhuizen toch een dagverplegingopname met als specialisme revalidatie geregistreerd. Deze records worden daarom uit de bestanden verwijderd (selectie: zorgtype = dagverpleging en ontslagspecialisme = revalidatie).

Gastverblijf

Gezonde personen die om één of andere reden in het ziekenhuis verblijven, moeten geregistreerd worden onder de code gastverblijf. In de meeste gevallen gaat het om gezonde zuigelingen die bij de opgenomen moeder moeten blijven, of omgekeerd. De scheidslijn tussen gezonde en ongezonde zuigeling of tussen gezonde en ongezonde moeder zal niet in alle gevallen eenduidig zijn en soms zullen wellicht gastverblijven toch als opname geregistreerd worden. Daarom is er voor gekozen om de gastverblijven in de bestanden te houden; bij de analyse van de data kunnen deze indien relevant, afhankelijk van het bestudeerde onderwerp, alsnog buiten beschouwing worden gelaten.

Onderhoudschemotherapie

Chemotherapie telt niet mee voor de adherentie bepaling. Ziekenhuizen dienen deze behandeling wel te registreren in de LMR, maar het vermoeden bestaat dat dit niet altijd compleet gebeurt. In de bestanden is te zien dat er toch behoorlijk veel onderhoudschemotherapie (diagnosecode V581) wordt opgegeven in dagverpleging: van ca. 44 000 in 1995 oplopend tot bijna 84 000 in 2001. De trend per ziekenhuis is verschillend. Naar alle waarschijnlijkheid is met deze gegevens een substantieel aandeel van alle chemotherapie geregistreerd. Daarom is er voor gekozen deze opnamen in de bestanden te houden. Bij publicatie van gegevens over deze specifieke behandeling zal vermeld moeten worden dat bij de gevonden aantallen waarschijnlijk sprake is van enige onderschatting.

3.3 Buitenlanders

Vóór de koppeling met de GBA worden de records in de LMR die behoren bij niet in Nederland wonende personen (selectie: postcode is blanco en landcode is niet 'NL') uit de bestanden verwijderd. Deze LMR-records van in Nederland behandelde buitenlanders worden niet meegenomen, omdat in CBS publicaties op niveau van de populatie uitgegaan wordt van de in de GBA geregistreerde bevolking.

In de LMR 1995–2001 bevat 0,31% van de records een landcode. Het gaat hier vrijwel uitsluitend om buitenlandse patiënten, die voor het merendeel (zes op de tien) afkomstig zijn uit België en Duitsland. Deze records van buitenlandse patiënten worden derhalve niet tot de basispopulatie gerekend en daarom uit de bestanden verwijderd. Normaliter wordt van de Nederlandse patiënten geen landcode geregistreerd. Er zijn slechts enkele records (n=711; <0,0001%) die een landcode 'NL' hebben. Dit zijn dus in het ziekenhuis opgenomen Nederlanders waarvan geen postcode geregistreerd is in de LMR: deze zullen om deze reden niet koppelen met de GBA maar behoren wel tot de basispopulatie van te koppelen records en worden dus niet vooraf uit de bestanden verwijderd.

Tabel 3
Percentage opnamen waarvan LMR-microgegevens bekend zijn naar type ziekenhuis

	Algemeen	Academisch	Categoriaal 1 ¹⁾	Categoriaal 2 ²⁾
1995	99,7	96,0	56,7	98,6
1996	99,5	97,7	58,6	98,9
1997	100,0	97,8	55,8	98,3
1998	99,8	97,1	68,9	98,3
1999	99,9	100,0	66,7	100,0
2000	99,6	100,0	57,2	87,7
2001	99,5	100,0	68,6	100,0

¹⁾ Categoriele ziekenhuizen die tot de 'landelijke pool' van Prismant behoren.

²⁾ Idem, maar exclusief de ziekenhuizen die in het geheel geen microgegevens hebben aangeleverd.

3.4 Gegeneerde records (missende microgegevens)

Prismant heeft onder ziekenhuisnummers 700–999 records gegeneerd in de LMR-bestanden om te completeren voor ziekenhuizen die hun gegevens niet, of niet volledig, hebben aangeleverd. Het gaat hier dus om *niet* in de LMR geregistreerde opnamen (missende records), waarvoor via dupliceren van bestaande records (genereren) is gecorrigeerd op basis van de wel bekende randtotalen. Prismant heeft deze missende records via genereren aangevuld om zodoende te komen tot een tabelleerbaar, op opnameniveau consistent gemaakt bestand, dat aansluit op de adherentiegegevens van de ziekenhuizen.

Aangezien het hier geen werkelijke opnamegegevens betreft dienen deze gegeneerde gegevens niet gekoppeld te worden aan personen in de GBA. De gegeneerde records, 0,6–1,5% van het totaal aantal LMR-records per jaar, worden daarom voorafgaande aan de koppeling uit de bestanden verwijderd.

Bij de analyse van de gekoppelde gegevens dient men zich bewust te zijn van deze missende microgegevens. Bij analyses op opnameniveau kan bijvoorbeeld door middel van weging voor de missende records gecorrigeerd worden. Bij de analyses op persoonsniveau ligt dit gecompliceerder en hangt het af van de vraagstelling of en hoe hiervoor gecorrigeerd kan worden. In ieder geval zal bij de analyse nagegaan dienen te worden wat de mogelijke invloed is van de missende records op de resultaten. Enkele aandachtspunten worden hieronder kort aangestipt.

Bij de missende records gaat het soms om (een gedeelte van) een bepaald soort opnamen dat ontbreekt (bijvoorbeeld dagopnamen of opnamen bij bepaalde afdelingen), soms mist een meer 'random' gedeelte van alle opnamen (bijvoorbeeld de laatste maand van een kalenderjaar) en bij een enkel ziekenhuis missen alle opnamen. De missende records hebben niet elk jaar betrekking op dezelfde ziekenhuizen. Bij sommige ziekenhuizen zijn de microgegevens voor bepaalde jaren wel compleet aanwezig en voor andere jaren niet of niet compleet. Hierdoor zou bij de statistische uitkomsten een stijging of daling van een bepaald soort opnamen te zien kunnen zijn, terwijl in werkelijkheid een ziekenhuis of afdeling werkelijke microgegevens over opnamen is gaan aanleveren waar ze dat voorheen niet deden of andersom. Op het totaal aantal opnamen zijn deze kleine verschillen meestal niet van belang. Maar wanneer de cijfers worden uitgesplitst naar bijvoorbeeld een bepaalde regio kunnen de verschillen soms wel zichtbaar zijn.

Wanneer men de populatie van ziekenhuizen dusdanig zou willen afbakenen dat dit soort verschillen in het geheel niet meer optreedt zou een substantieel deel van de gegevens van Prismant

ongebruikt blijven. Omdat de totale verschillen in de tijd klein zijn lijkt het beter om zoveel mogelijk informatie te gebruiken en waar relevant de verschillen in ziekenhuispopulaties aan te geven in tekst of voetnoten. Bovendien is het zo dat wanneer de populatie wordt afgebakend volgens criteria waarmee de inconsistenties van niet geregistreerde records buiten beeld blijven, er weer andere inconsistenties naar voren komen. Als bijvoorbeeld uitsluitend compleet registrerende algemene of academische ziekenhuizen geselecteerd zouden worden, ontstaan er toch inconsistenties in de populatie wanneer categorale instellingen na fusie opgaan in algemene of academische ziekenhuizen.

In tabel 3 wordt per jaar aangegeven van hoeveel procent van de opnamen er werkelijke microgegevens in de LMR aanwezig zijn. Categoriaal 1 geeft het percentage voor alle tot de 'landelijke pool' van Prismant behorende categorale ziekenhuizen (zie 3.1) en categoriaal 2 geeft het percentage voor dezelfde ziekenhuizen met uitzondering van de 1 à 2 kankerklinieken die in het geheel geen microgegevens hebben aangeleverd.

Uit tabel 3 blijkt dat de dekking van de categorale ziekenhuizen exclusief de twee kankerklinieken redelijk groot is, de dekking van de algemene ziekenhuizen nagenoeg compleet is, en bij de academische ziekenhuizen vanaf 1999 een complete deelname is. Of de missende records wel of niet van enig belang zijn, hangt af van de analysevraagstelling, bijvoorbeeld welk type opnamen, welke diagnose/specialisme, wel of niet regionaal, welke jaren. De onderzoeker kan dan met behulp van de beschikbare informatie over de missende records (soort ziekenhuis, soort opname, specialisme, postcode patiënt), desgewenst een schatting maken van de (maximale) invloed die de missende gegevens op de resultaten kunnen hebben.

Ten slotte dient hier nog opgemerkt te worden dat opnamen in buitenlandse ziekenhuizen van in de GBA ingeschreven personen worden gemist, omdat de LMR alleen betrekking heeft op Nederlandse ziekenhuizen. Hier kan dan ook niet voor gecorrigeerd worden, tenzij hierover informatie uit andere registraties beschikbaar komt.

3.5 Overzicht afbakening LMR-bestanden

In tabel 4 wordt een cijfermatig overzicht gegeven van de afbakening van de LMR-bestanden zoals in dit hoofdstuk beschreven. Per LMR-bestandsjaar wordt aangegeven hoeveel records uit de bestanden zijn verwijderd en om welke reden, in de volgorde zoals in de tabel beschreven.

Tabel 4
Afbakening van de LMR-bestanden (1995–2001) vóór koppeling met de GBA

	1995	1996	1997	1998	1999	2000	2001	Totaal	
								<i>abs.</i>	<i>%</i>
Oorspronkelijk aantal records in LMR	2 236 410	2 289 395	2 326 404	2 348 414	2 360 475	2 368 194	2 454 115	16 383 407	100,00
<i>Exclusie van records betreffende:</i>									
ontbrekend registratiejaar en instellingsnummer ¹⁾	–	2	1	2	–	–	–	5	0,00
onterecht geregistreerde behandelingen	5 549	6 717	7 010	8 247	6 205	6 113	6 161	46 002	0,28
buitenlandse patiënten	7 134	7 112	7 448	7 106	7 436	7 233	6 995	50 464	0,31
gegenereerde records	33 082	31 327	22 129	22 011	15 055	21 230	21 387	166 221	1,01
astmacentra	805	584	638	697	782	784	813	5 103	0,03
Afgebakende LMR-populatie vóór koppeling met de GBA	2 189 840	2 243 653	2 289 178	2 310 351	2 330 997	2 332 834	2 418 759	16 115 612	98,37

¹⁾ Door het ontbreken van deze variabelen kunnen deze records niet gelinkt worden aan de LMR-deelbestanden met gegevens over diagnoses, verrichtingen en specialismen. Vanwege de daarom zeer beperkte gebruikwaarde en het zeer lage aantal zijn deze records ook uit de bestanden verwijderd.

4. Koppeling van LMR aan GBA-gegevens

In dit hoofdstuk worden de gevolgde procedure voor de koppeling van de LMR met de GBA beschreven en de resultaten daarvan voor de jaren 1995–2001. Bij deze koppeling is gebruik gemaakt van de koppelsleutel postcode (6-posities wanneer beschikbaar in de LMR, en anders alleen op de 4-cijferige postcode), geboortedatum en geslacht. Bij de meervoudig-gekoppelden is ook nog gekoppeld op datum van overlijden, in geval van in het ziekenhuis overleden LMR-patiënten.

Daarnaast is een koppeling uitgevoerd op alleen de 4-cijferige postcode, geboortedatum en geslacht, omdat dit voor sommige toepassingen van nut kan zijn.

Bij de hier uitgevoerde koppelingen is geen gebruik gemaakt van het patiëntnummer in de LMR. Aangezien deze variabele niet in de GBA voorkomt kan deze niet gebruikt worden als primaire koppelvrijvariabele. Wel kan het patiëntnummer in principe gebruikt worden om niet uniek gekoppelde records 'bij te koppelen' bij wel uniek gekoppelde records. In hoofdstuk 7 wordt verder op deze mogelijkheid en de beperkingen daarvan ingegaan.

4.1 Het CKP

De koppeling met de GBA is uitgevoerd met het Centraal Koppelbestand Personen (CKP), een longitudinaal bestand van personen die ingeschreven staan of hebben gestaan in de GBA (zie ook 2.2). Het CKP-bestand begint op 1 januari 1995. Het longitudinale karakter van het bestand houdt in dat er van één persoon meerdere records voorkomen als zich wijzigingen in de persoonsgegevens (bijv. burgerlijke staat, nationaliteit, emigratie) en/of in het inschrijvingsadres (verhuizingen, postcode wijzigingen) hebben voorgedaan.

De GBA bevat een uniek persoonsidentificatienummer, het A-nummer. In het CKP is dit nummer vervangen door een betekenisloos nummer 'RINPersoon'. Iedere mutatie wordt vastgelegd door aan RINPersoon een volgnummer toe te kennen. Ieder record is bovendien voorzien van een geldigheidsperiode en een mutatiecode die aangeeft waarom het desbetreffende record is beëindigd of gestart (bijv. als gevolg van adreswijziging). De mutatiecodes staan vermeld in Bijlage 2. De voor de koppeling gebruikte versie van het CKP omvat alle personen die in de periode 1995/01/01 t/m 2002/01/01 in de GBA ingeschreven hebben gestaan.

4.2 Voorbewerking LMR voor koppeling met CKP

De aan het CBS geleverde LMR-gegevens zijn onderverdeeld in een aantal bestanden: de zogenaamde 'oposbestanden' met de algemene gegevens van patiënt, opname en ontslag; de diagnosesbestanden met daarin gegevens over diagnoses en overdracht per verantwoordelijkheidsperiode binnen de opname; de verrichtingenbestanden met daarin de tijdens de opname uitgevoerde verrichtingen (met dezelfde verantwoordelijkheidsperiode systemiek) en de specialismenbestanden met de betrokken specialismen per verantwoordingsperiode. De gegevens in de verschillende bestanden zijn onderling koppelbaar op registratiejaar, instellingsnummer, opnamenummer en, wanneer relevant, verantwoordelijkheidsperiodenummer. In Bijlage 1 staan de in de bestanden opgenomen variabelen vermeld.

Voor de koppeling zijn de LMR-records uit de oposbestanden van de jaren 1995–2001 samengevoegd in één bestand. Verder is uit het diagnosesbestand de overdrachtsdatum van de laatste verantwoordelijkheidsperiode ('indicatie overdracht' is dan 'ontslag of overlijden') toegevoegd aan het desbetreffende opnamerecord in het oposbestand. In combinatie met de variabele 'bestemming' in het oposbestand, waarin wordt aangegeven 'in instelling overleden' bij overlijden van de patiënt tijdens de opname, zijn hierdoor in de oposbestanden zowel de opnamedatum als de ontslagdatum en het eventueel overlijden (ontslagdatum is dan tevens overlijdensdatum) in één record samengebracht.

4.3 Het koppelen van LMR aan CKP

Koppelingprocedure

Bij het koppelen van de LMR-gegevens aan het CKP is de systematiek gevolgd die het CBS meestal voor dergelijke koppelingen hanteert.

De koppeling vindt plaats door op of rond een bepaalde datum (hier: de *opnamedatum* van het LMR-record) de dan geldige CKP-records te selecteren die overeenkomen op de koppelsleutel. Omdat zowel bij de GBA als bij de LMR administratieve verschillen kunnen optreden bij het registreren van verhuisdata, wordt een zekere marge rond de opnamedatum toegelaten. Testen zijn uitgevoerd met verschillende marges rond de opnamedatum: +/- 0, 15, 30, 90 en 180 dagen. Omdat de LMR een ontslagregistratie is – waarin dus de in het desbetreffende kalenderjaar uit het ziekenhuis *ontslagen* patiënten zijn geregistreerd – komen er in de LMR 1995–2001 ook opnamedata van vóór 1995 voor (met ontslag op of na 1 januari 1995) en ontbreken de patiënten die vóór 2002 zijn opgenomen maar pas op of na 1 januari 2002 zijn ontslagen. Daar bij de koppeling de opnamedatum gehanteerd is en het gebruikte CKP-bestand alleen gegevens bevat vanaf 1 januari 1995 t/m 1 januari 2002 heeft dit een paar gevolgen. Wanneer een marge van 30 dagen rond de opnamedatum gehanteerd wordt zijn de LMR-records met een opnamedatum vóór 1 januari 1995, ten dele niet koppelbaar als de opnamedatum op 1 december 1994 of eerder valt (0,2% van de LMR-ontslagrecords uit 1995), en ten dele alleen koppelbaar met de gehanteerde plusmarge rond de opnamedatum, nl. als de opnamedatum tussen 1 december 1994 en 1 januari 1995 valt (1,1% van de LMR-records uit 1995). Verder kan met deze marge van 30 dagen bij de LMR-opnamen met een opnamedatum in december 2001 niet altijd de volledige plusmarge rond de opnamedatum gehanteerd worden.

Als er n LMR-records met m CKP-records gekoppeld kunnen worden (dus de n en m records hebben allen dezelfde waarde op de koppelsleutel), wordt ieder LMR-record m maal gekopieerd en worden de koppelbare CKP-gegevens toegevoegd. Het totaal aantal records na de koppeling wordt dus $n \times m$. Er kunnen 0, 1 of >1 (m) CKP-records gekoppeld worden aan 1 of >1 (n) LMR-records.

Bij het bepalen van de m koppelbare CKP-records wordt rekening gehouden met het feit dat een persoon in meer dan één CKP-record vertegenwoordigd kan zijn. Als er meer dan één record van dezelfde persoon voor koppeling in aanmerking komt – doordat er binnen de gekozen marge rond de opnamedatum mutaties in de demografische gegevens van deze persoon hebben plaatsgevonden – wordt het CKP-record met de kleinste tijdsafstand tot de opnamedatum gekozen. Op deze wijze wordt ervoor gezorgd dat

er maar 1 CKP-record per GBA-persoon aan een LMR-record gekoppeld kan worden.

Bij een succesvolle koppeling, dat wil zeggen als er 1 of meer CKP-records worden gekoppeld met een LMR-record, worden er verschillende gegevens aan het LMR-record toegevoegd, zoals de RINPersoon, het (betekenisloze) persoonsidentificatienummer. Daarnaast wordt onder andere ook een variabele toegevoegd die de 'afstand' aangeeft van de opnamedatum tot de geldigheidsperiode van het aangekoppelde CKP-record. De maximale afstand is gelijk aan de marge rond de opnamedatum die bij de koppeling is gehanteerd. Tevens worden tellervariabelen toegevoegd die inzicht geven in de n x m koppeling, zoals het aantal aangekoppelde CKP-records. Aangezien het koppelingsproces in verschillende stappen verloopt, is ook een variabele toegevoegd die aangeeft welke koppelingsstappen het desbetreffende LMR-record heeft doorlopen. In Bijlage 2 is een volledig overzicht gegeven van de bij de koppeling toegevoegde gegevens.

Uitgevoerde koppelingsstappen

Bij de koppeling van de LMR aan het CKP zijn de volgende stappen uitgevoerd:[†]

- I. Koppeling van LMR op *postcode (6 posities), geboortedatum, geslacht en opnamedatum*, hetgeen oplevert:
 - Ia: gekoppeld aan één CKP-record.
 - Ib: gekoppeld aan meer dan één CKP-record: n x m records.
 - Ic: niet gekoppeld aan CKP.
- II. Koppeling van Ic op *postcode (4 posities), geboortedatum, geslacht en opnamedatum*, hetgeen oplevert:
 - Ila: gekoppeld aan één CKP-record.
 - Ilb: gekoppeld aan meer dan één CKP-record: n x m records.
 - Ilc: niet gekoppeld aan CKP.

Na bovenstaande twee koppelstappen zijn er drie bestanden ontstaan:

- de LMR-records die gekoppeld zijn aan één CKP-record, bestaande uit Ia + Ila.
- de LMR-records die gekoppeld zijn aan meer dan één CKP-record, bestaande uit Ib + Ilb.
- de niet aan het CKP gekoppelde LMR-records, bestaande uit Ic.

Vervolgens zijn de meervoudig gekoppelde LMR-records (Ib + Ilb) waarbij sprake is van overlijden in het ziekenhuis, nog gekoppeld op datum van overlijden:[‡]

III. Koppeling van Ib + Ilb op *overlijdensdatum*, waarbij sterfdatum GBA = sterfdatum LMR +/- 1 dag (marge).

Na stap III zijn een aantal records van de eerder meervoudig gekoppelde records naar het bestand van uniek gekoppelde records gegaan, hetgeen uiteindelijk resulteert in de volgende eindbestanden:

1. uniek gekoppelde LMR-records, bestaande uit Ia + Ila + uniek-gekoppelden in III.
2. meervoudig gekoppelde LMR-records, bestaande uit Ib + Ilb – uniek-gekoppelden in III.
3. niet gekoppelde LMR-records, bestaande uit Ic.

De verschillende koppelingsstappen zijn in figuur 1 nog eens grafisch weergegeven.

Er is bij de koppeling van afgezien om koppelwinst na te streven door bij de meervoudig-gekoppelden van stap II RINpersonen uit te sluiten die al in stap I gekoppeld zijn. In hoofdstuk 6 wordt dit nader beargumenteerd.

Omdat het voor sommige analyse-toepassingen de voorkeur kan hebben om alleen een koppeling op basis van de 4-cijferige postcode (en geboortedatum en geslacht) te gebruiken, is deze koppeling ook uitgevoerd. Dit is gedaan door de uniek-gekoppelden van stap I nogmaals te koppelen op de sleutel met 4-cijferige postcode en de resulterende uniek-gekoppelden op te tellen bij die van stap II en de meervoudig-gekoppelden bij die van stap I en II. De niet-gekoppelden zijn gelijk aan die van stap II. Deze procedure is gelijk aan een aparte koppeling van het hele bestand op de postcode-4 sleutel. Door deze koppeling op 4-cijferige postcode te vergelijken met die op 6 posities kan ook bestudeerd worden wat het effect is van het al of niet ter beschikking hebben van de 2 letters van de postcode.

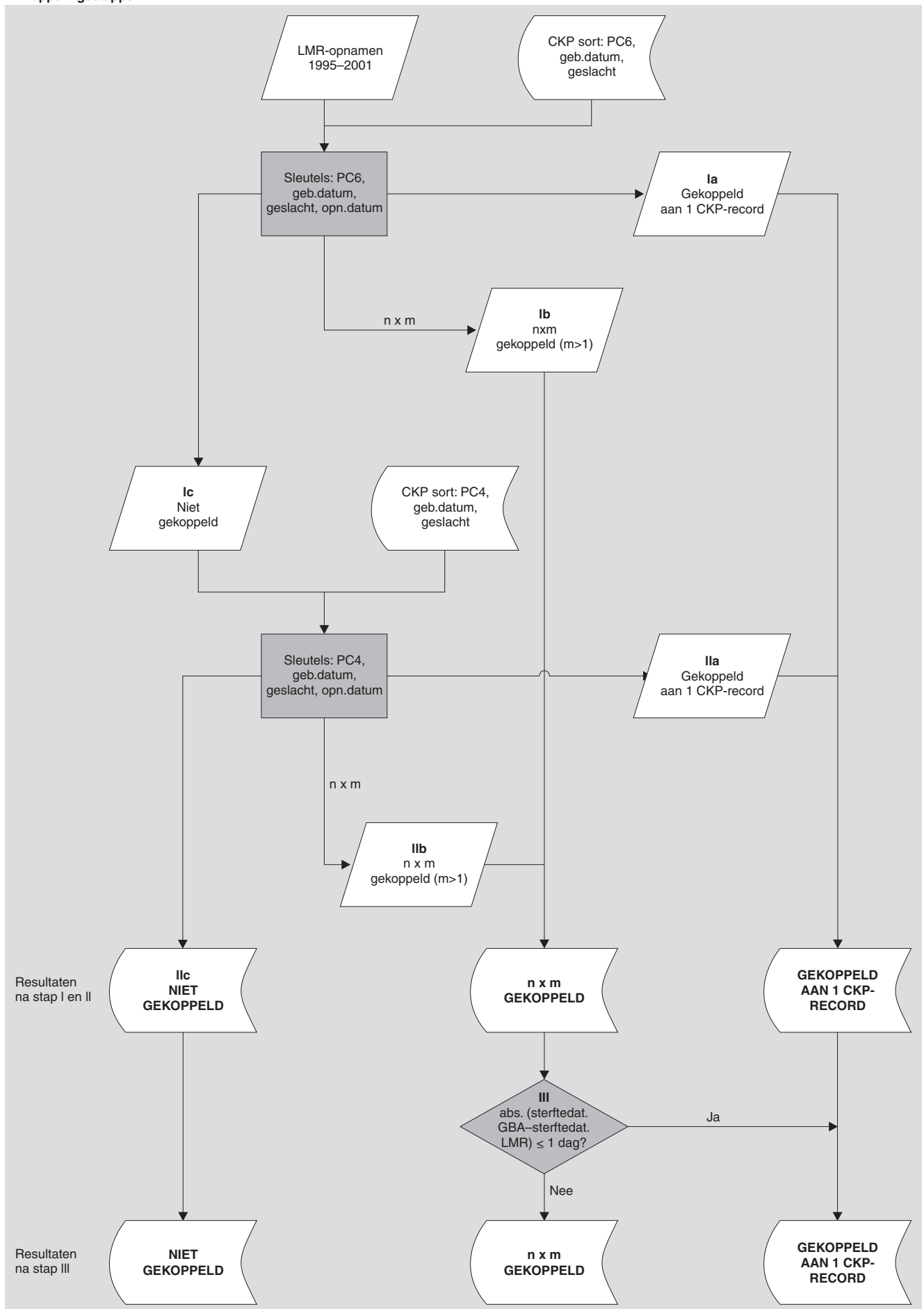
[†] Koppelingsstappen I en II zijn uitgevoerd met verschillende marges rond de opnamedatum: +/- 0, 15, 30, 90 en 180 dagen.

[‡] Koppelingsstap III is alleen uitgevoerd op de resultaten met een marge van 30 dagen rond de opnamedatum en een marge van 1 dag rond de overlijdensdatum.

Tabel 5
Rendementen koppeling LMR met GBA na koppelingsstappen I en II, bij verschillende marges rond de opnamedatum

	Gehanteerde marge rond opnamedatum									
	+/- 0 dagen		+/- 15 dagen		+/- 30 dagen		+/- 90 dagen		+/- 180 dagen	
	abs.	%	abs.	%	abs.	%	abs.	%	abs.	%
Totaal aantal LMR-records (1995–2001)	16 115 612	100,0	16 115 612	100,0	16 115 612	100,0	16 115 612	100,0	16 115 612	100,0
Stap I Kopp. op postc. 6 pos.+gebdat.+sexe+opnamedat.										
Ia. gekoppeld aan 1 GBA-persoon	4 062 266	25,2	4 077 250	25,3	4 086 747	25,4	4 108 989	25,5	4 124 160	25,6
Ib. gekoppeld aan >1 GBA-persoon	28 750	0,2	28 983	0,2	29 172	0,2	29 716	0,2	30 360	0,2
Ic. niet gekoppeld (naar stap II)	12 024 596	74,6	12 009 379	74,5	11 999 693	74,5	11 976 907	74,3	11 961 092	74,2
Stap II Kopp. van Ic op postc. 4 pos.+gebdat.+sexe+opnamedat.										
Ila. gekoppeld aan 1 GBA-persoon	9 991 674	62,0	10 011 162	62,1	10 018 374	62,2	10 023 093	62,2	10 005 063	62,1
Ilb. gekoppeld aan >1 GBA-persoon	1 372 802	8,5	1 384 457	8,6	1 394 213	8,7	1 428 498	8,9	1 474 579	9,2
Ilc. niet gekoppeld	660 120	4,1	613 760	3,8	587 106	3,6	525 316	3,3	481 450	3,0
Resultaat na Stap I en Stap II										
gekoppeld aan 1 GBA-persoon (uniek gekoppeld)	14 053 940	87,2	14 088 412	87,4	14 105 121	87,5	14 132 082	87,7	14 129 223	87,7
gekoppeld aan >1 GBA-persoon (meervoudig gekoppeld)	1 401 552	8,7	1 413 440	8,8	1 423 385	8,8	1 458 214	9,0	1 504 939	9,3
niet gekoppeld	660 120	4,1	613 760	3,8	587 106	3,6	525 316	3,3	481 450	3,0

1. Koppelingsstappen



Tabel 6
Resultaten koppeling LMR met GBA na koppelingsstap I, II en III

	Abs.	%
Totaal aantal LMR-records (1995–2001)	16 115 612	100,0
Stap I Kopp. op postc. 6 pos.+gebdat.+sexe+opnamedat. ¹⁾		
la. gekoppeld aan 1 GBA-persoon	4 086 747	25,4
lb. gekoppeld aan >1 GBA-persoon	29 172	0,2
lc. niet gekoppeld (naar stap II)	11 999 693	74,5
Stap II Kopp. van lc op postc. 4 pos.+gebdat.+sexe+opnamedat. ¹⁾		
IIa. gekoppeld aan 1 GBA-persoon	10 018 374	62,2
IIb. gekoppeld aan >1 GBA-persoon	1 394 213	8,7
IIc. niet gekoppeld	587 106	3,6
Resultaat na Stap I en Stap II		
gekoppeld aan 1 GBA-persoon (uniek gekoppeld)	14 105 121	87,5
gekoppeld aan >1 GBA-persoon (meervoudig gekoppeld)	1 423 385	8,8
niet gekoppeld	587 106	3,6
Stap III Kopp. van lb en IIb op sterfededat. GBA = sterfededat. LMR +/- 1 dag		
Resultaat na Stap III		
gekoppeld aan 1 GBA-persoon (uniek gekoppeld)	14 121 787	87,6
gekoppeld aan >1 GBA-persoon (meervoudig gekoppeld)	1 406 719	8,7
niet gekoppeld	587 106	3,6

¹⁾ Met marge van +/- 30 dagen.

4.4 Resultaten na koppelingsstappen I en II

In tabel 5 zijn de resultaten weergegeven van de koppelingsstappen I en II, uitgaande van verschillende marges rond de opnamedatum.

Uit tabel 5 blijkt dat het overgrote deel van de LMR-records op de opnamedatum zelf koppelt; het toelaten van een marge levert 0,2–0,5% extra unieke koppelingen op. Naarmate de toegelaten marge groter wordt, neemt het aantal meervoudig-gekoppelden logischerwijs toe, en het aantal niet-gekoppelden af. Bij het bepalen van de gewenste marge moet een afweging gemaakt worden tussen de ongewenste toename van de meervoudig-gekoppelden en de gewenste afname van de niet-gekoppelden. Uit tabel 5 blijkt dat bij een marge van +/- 30 dagen een geschikte 'trade-off' bereikt wordt: het percentage meervoudig-gekoppelden is met slechts 0,1 toegenomen, terwijl het percentage niet-gekoppelden met 0,4% is afgenomen. Op grond hiervan is besloten om bij de definitieve koppeling een marge van 30 dagen aan te houden.

Het relatief lage percentage unieke koppelingen (ten opzichte van het totaal aantal records) bij stap I is een logisch gevolg van het feit dat slechts bij ongeveer een kwart van de LMR-records de 6 posities van de postcode gevuld was, van de overige records is alleen de 4-cijferige postcode bekend. De niet gekoppelde records

bij stap I bestaan dan ook voornamelijk (97,4%) uit records die alleen een 4-cijferige postcode hebben; deze gaan door naar stap II. Van alle records met een postcode van 6 posities (n=4 426 463) koppelt uiteindelijk, na stap I en stap II en bij een gehanteerde marge van +/- 30 dagen, 95,1% uniek, 1,1% meervoudig en 3,8% koppelt niet. Als alle ziekenhuizen een postcode-6 zouden registreren in de LMR zou dus naar verwachting een koppelement van ca. 95% gehaald kunnen worden.

4.5 Resultaten na koppelingsstap III

De laatste koppelingsstap is alleen uitgevoerd op de eerdere koppelingsresultaten bij een gehanteerde marge van +/- 30 dagen rond opnamedatum. In tabel 6 worden hiervan de resultaten gepresenteerd.

Uit tabel 6 blijkt dat de extra koppelingsstap op overlijdensdatum tot een relatief gering aantal extra unieke koppelingen leidt (0,1%) ten opzichte van het resultaat na de stappen I en II. De meeste van de in het ziekenhuis overleden patiënten zullen al in deze eerste twee stappen gekoppeld zijn. De toename van 0,1% koppelingen komt overigens ongeveer overeen met wat theoretisch verwacht kan worden: uitgaande van 1% overlijdensgevallen in de

Tabel 7
Resultaten koppeling LMR met GBA na koppelingsstap III, per jaar

	Totaal aantal LMR-records		Gekoppeld aan 1 GBA-persoon		Gekoppeld aan > 1 GBA-persoon		Niet gekoppeld	
	abs.	%	abs.	%	abs.	%	abs.	%
1995	2 189 840	100,0	1 907 240	87,1	192 414	8,8	90 186	4,1
1996	2 243 653	100,0	1 963 191	87,5	193 605	8,6	86 857	3,9
1997	2 289 178	100,0	2 006 674	87,7	197 624	8,6	84 880	3,7
1998	2 310 351	100,0	2 023 726	87,6	201 927	8,7	84 698	3,7
1999	2 330 997	100,0	2 046 487	87,8	203 170	8,7	81 340	3,5
2000	2 332 834	100,0	2 048 326	87,8	204 857	8,8	79 651	3,4
2001	2 418 759	100,0	2 126 143	87,9	213 122	8,8	79 494	3,3
1995–2001	16 115 612	100,0	14 121 787	87,6	1 406 719	8,7	587 106	3,6

bevolking per jaar, zouden grofweg 1% van de 8,8% (≈ 0,09%) meervoudig-gekoppelden na stap II alsnog gekoppeld kunnen worden op overlijdensdatum.

Het eindresultaat na de drie koppelingsstappen is dat 87,6% van de LMR-records uniek gekoppeld kon worden aan de GBA, bij 8,7% was sprake van een meervoudige koppeling en 3,6% van de records kon niet gekoppeld worden.

Dit koppelingsresultaat kan bevredigend genoemd worden. Aan gezien zowel de GBA als de LMR onderhevig zijn aan een administratieve foutenruis – die bij de GBA wat het adresgegeven betreft tussen de 1 en 2% ligt (zie 2.2); bij de LMR is deze onbekend – is een percentage van 3,6% niet-gekoppelde records zeer redelijk. Temeer daar de LMR behalve opnamen van personen die op een ander adres in de GBA staan ingeschreven (bijvoorbeeld sommige studenten, daklozen) ook opnamen zal bevatten van in Nederland wonende personen die niet (meer) in de GBA staan ingeschreven. Bij dit laatste kan het bijvoorbeeld gaan om tijdelijk in Nederland verblijvende personen die vanwege hun korte verblijf niet in de GBA zijn ingeschreven (ook: geëmigreerden die voor een ziekenhuisopname terug naar Nederland komen en daarbij een adres in Nederland opgeven) of illegalen.[§]

De 8,7% meervoudig gekoppelde records zijn een natuurlijk gevolg van het beperkte discriminerende vermogen van de koppelsleutel, met name bij de records met alleen een viercijferige postcode.

In tabel 7 zijn de eindresultaten van de koppeling na stap III uitgesplitst naar jaar van de LMR.

Hieruit blijkt dat de koppelingsresultaten over de verschillende jaren vrij stabiel zijn. Wel is te zien dat het percentage niet gekoppelde records geleidelijk afneemt van 4,1% in 1995 tot 3,3% in 2001. Het hogere percentage in 1995 komt voor een deel omdat 0,2% van de opnamen met ontslag in 1995 niet koppelen omdat zij een opnamedatum hebben van 1 december 1994 of eerder: als deze records niet worden meegenomen dan is het percentage niet gekoppelde records in 1995 gelijk aan dat van 1996 (3,9%).

4.6 Resultaten van koppeling op alleen postcode-4 sleutel

In tabel 8 zijn de resultaten weergegeven van de koppeling op alleen 4-cijferige postcode, geboortedatum en geslacht, waarbij eveneens een marge van +/- 30 dagen rond de opnamedatum gehanteerd is.

Deze basale koppeling geeft 84,1% uniek-gekoppelden: een verschil van 3,4 procentpunt vergeleken met de resultaten na stap I en II. De meervoudig-gekoppelden zijn met ditzelfde verschilpercentage verhoogd: dit zijn dus de records die op postcode-6 wel uniek koppelen en op postcode-4 niet. Het aantal niet-gekoppelden blijft vanzelfsprekend gelijk.

[§] De in de LMR als buitenlander geregistreerde personen zijn wel al voor de koppeling uit de bestanden verwijderd.

Tabel 8
Resultaten koppeling LMR met GBA op alleen 4-cijferige postcode, geboortedatum en geslacht

	Abs.	%
Totaal aantal LMR-records (1995–2001)	16 115 612	100,0
Koppeling op postcode 4 pos.+gebdat.+sexe+opnamedat. ¹⁾		
gekoppeld aan 1 GBA-persoon (uniek gekoppeld)	13 561 014	84,1
gekoppeld aan >1 GBA-persoon (meervoudig gekoppeld)	1 967 492	12,2
niet gekoppeld	587 106	3,6

¹⁾ Met marge van +/- 30 dagen.

5. Onderzoek naar de kwaliteit van de koppeling

In de vorige paragrafen is de kwaliteit van het koppelingsresultaat al in algemene zin beschreven door aan te geven dat de gevonden percentages uniek, meervoudig en niet gekoppelde records plausibel zijn. In dit hoofdstuk wordt meer in detail ingegaan op de kwaliteit van de koppeling door foutenbronnen in de koppelvariabelen aan te geven, de uniek gekoppelde records te checken op foutieve koppelingen en de selectiviteit van de niet gekoppelde records te beschrijven.

5.1 De koppelvariabelen in de LMR

De LMR-bestanden zijn eerst nagegaan op bijzonderheden ten aanzien van de koppelvariabelen. Dit kan namelijk van invloed zijn op de koppelingsresultaten.

Geboortedatum

De geboortedata zijn 'op het oog' gecheckt op uitbijters en onwaarschijnlijkheden. De uitbijters betreffen vooral de geboortedagen 1 januari en 1 juli. Dit is een bekend fenomeen: deze dagen worden namelijk vaak ingevuld als de echte geboortedatum niet bekend is. Verder komen er geboortedata van vóór 1900 voor in de LMR-bestanden. De aantallen per geboortjaar zijn echter plausibel. Steeds meer mensen in Nederland worden ouder dan 100 en dus kunnen geboortedata van voor 1900 voorkomen in de bestanden over 1995–2001. Verder bevat het LMR-bestand van 1995 één ziekenhuis met 218 records met een ongeldige geboortedatum: 01010019. Deze records zijn hier niet op gecorrigeerd, omdat onduidelijk is wat de werkelijke geboortedata van deze patiënten zijn. Bij de koppeling zullen deze records in de niet-gekoppelde groep terechtkomen.

Postcode

Om inzicht te krijgen in de kwaliteit van het postcode-gegeven zijn de in de LMR voorkomende postcodes vergeleken met de postcodes in de GBA en in een postcodebestand afkomstig van TPG Post. Gebruikt zijn de TPG-postcodes per mei 2003, hetgeen als consequentie heeft dat enkele oude postcodes die voor deze datum zijn opgeheven, maar ten dele nog wel actueel konden zijn in de jaren 1995–2001, niet in het bestand voorkomen. Als gevolg hiervan kan het aantal LMR-postcodes dat niet in het gebruikte TPG-bestand teruggevonden kon worden een lichte overschatting zijn van de werkelijk niet bestaande TPG-postcodes. Verder wordt opgemerkt dat de postcodes in de GBA alleen de postcodes

behorend bij een bewoond adres en behorend bij een in de GBA ingeschreven persoon betreffen; het is geen complete lijst van alle voorkomende postcodes.

In tabel 9 en 10 zijn resultaten van de vergelijking weergegeven.

Uit tabel 9 blijkt dat over alle jaren 1995–2001 van 0,2% van de LMR-records de postcode-4 niet voorkomt in de GBA. Deze records zullen dan ook niet met de GBA koppelen. Hierbij zit ook een klein aantal records (ca. 700) met blanco postcode.

Van de in totaal 4 426 463 LMR-records met een postcode van 6 posities, zijn bij 4 416 266 records ook inderdaad 2 letters op de laatste 2 posities ingevuld. Dit zijn dus potentieel geldige postcode-6 gegevens. Bij de overige 10 197 (0,2%) is er wat anders ingevuld op de laatste 2 posities. Dit kan voorkomen omdat de deelnemende ziekenhuizen in principe vrij zijn hoe zij deze 2 posities vullen; dit wordt ook niet gecontroleerd door Prismant. Deze postcodes zijn logischerwijs noch in de GBA, noch in het TPG referentiebestand te vinden. De desbetreffende records zullen niet koppelen op postcode-6 (stap I), maar wel naar de postcode-4 koppeling (stap II) doorgaan. In tabel 9 zijn deze 10 197 records niet bij de postcode-6 groep, maar bij de postcode-4 groep geplaatst en dus alleen op grond van de postcode-4 met het TPG-bestand vergeleken.

Uit tabel 9 blijkt dat bij 0,8% van de LMR-records met een potentieel geldige postcode-6 de desbetreffende postcode niet voorkomt in de GBA. Bij 0,5% komt deze ook niet in het TPG-referentiebestand voor, hetgeen meestal wijst op invoerfouten bij de LMR.

In tegenstelling tot de wijkcode (de laatste 2 posities van de postcode) is de viercijferige postcode een verplichte variabele in de LMR, die ook door Prismant wordt gecontroleerd. Uit tabel 9 blijkt dat er inderdaad maar weinig records zijn met een postcode-4 die noch in de GBA noch in het TPG-referentiebestand voorkomen (0,01%). Bij deze records is er waarschijnlijk ook sprake van invoerfouten bij de LMR.

Berekend op alle LMR-records komt 0,4% van de daarin voorkomende postcodes, op 4 dan wel 6 posities, niet voor in de GBA.

De LMR-postcodes die niet in de GBA voorkomen maar wel in het TPG-referentiebestand zijn nader bekeken op de aard van de postcode, namelijk of de postcode naar een straatnaam, naar een postbus of naar beide verwijst. De resultaten zijn weergegeven in tabel 10.

Tabel 9
Postcodes in LMR (1995–2001) vergeleken met GBA en TPG-referentiebestand

	Postcode-gegeven					
	Wel in LMR Niet in GBA		Wel in LMR Niet in GBA Niet in TPG-ref.		Totaal in LMR	
	abs.	%	abs.	%	abs.	%
Postcodes in LMR						
postcode 4 posities	754	15,9	56	1,2	4 728	100,0
postcode 6 posities	17 082	6,3	13 184	4,9	270 429	100,0
LMR-records met						
postcode 4 posities	22 694	0,2	1 074	0,01	11 699 346	100,0
postcode 6 posities	36 011	0,8	24 204	0,5	4 416 266	100,0

Tabel 10
LMR-postcodes (1995–2001) die niet in de GBA voorkomen, maar wel in TPG-referentiebestand, naar aard van de postcode

	Straatnaam		Postbus		Straatnaam & postbus		Totaal in LMR	
	abs.	%	abs.	%	abs.	%	abs.	%
Postcodes in LMR								
postcode 4 posities	13	1,9	682	97,7	3	0,4	698	100
postcode 6 posities	2 140	54,9	1 758	45,1	–	–	3 898	100
LMR-records met								
postcode 4 posities	44	0,2	21 460	99,3	116	0,5	21 620	100
postcode 6 posities	4 923	41,7	6 884	58,3	–	–	11 807	100

Bij de LMR-records met een postcode-4 gaat het hier vrijwel uitsluitend om postbussen en bij de records met een postcode-6 om 58% postbussen en 42% straatnamen. Omdat deze niet in de GBA voorkomen gaat het hier waarschijnlijk voor een deel om adressen die geen woonadres zijn, zoals bedrijvenparken, hotels e.d. Daarnaast kan het gaan om adressen van personen die niet of niet op het juiste adres in de GBA staan ingeschreven.

Samengevat kan geconcludeerd worden dat maar een beperkt deel van de LMR-records niet koppelen met de GBA omdat het postcode-4 of postcode-6 gegeven niet voorkomt in de GBA (0,2% en 1,0% (0,8 + 0,2) van de LMR-records met een postcode-4, resp. postcode-6 gegeven). Dit wordt veroorzaakt door niet aan de postcode gerelateerde codes in de laatste 2 posities van de postcode, invoerfouten in de LMR en waarschijnlijk door LMR-postadressen die niet bij de GBA als (woon- of brief-) adres zijn opgegeven, en door patiënten die niet of niet op het juiste adres in de GBA staan ingeschreven.

Overlijdensdatum

Ten slotte is er ten aanzien van de overlijdensdatum nog nagegaan of er systematische afwijkingen zijn in de weekdag van overlijden zoals geregistreerd in de LMR en zoals geregistreerd in de GBA. Bij overlijden op weekenddagen zou er bijvoorbeeld een verschil kunnen zijn in aangifte datum bij de GBA en overlijdensdatum zoals geregistreerd in de LMR; dit zou dan van invloed kunnen zijn op de koppeling op overlijdensdatum.

Om dit te onderzoeken zijn alleen de in een ziekenhuis overleden GBA-persoon geselecteerd door gebruik te maken van de doodsoorzaakbestanden, waaruit de in een ziekenhuis overledenen (excl. psychiatrische ziekenhuizen, want deze nemen niet deel aan de LMR) kunnen worden afgeleid. De datum van overlijden in de doodsoorzakenbestanden is afkomstig uit de GBA. Uit de vergelijking bleek dat de procentuele verdeling van het overlijden naar weekdag (van zondag t/m zaterdag) vrijwel gelijk is bij de LMR- en de GBA/doodsoorzakenbestanden. Er is dus wat dit betreft geen systematisch verschil tussen deze registraties, dat van invloed zou kunnen zijn op de koppeling.

5.2 De koppelvariabelen in de GBA

Geboortedatum

In de GBA zijn bij de allochtone populatie de geboortedagen 1 januari en 1 juli duidelijk oververtegenwoordigd, met ruim 3, resp. 2 maal zoveel ingeschrevenen als op 'gewone' dagen. Ook de eerste van de andere maanden komt meer dan gemiddeld voor, maar in veel mindere mate dan de 1 januari en 1 juli uit-

schieters.⁷⁾ Bij de autochtone populatie zijn deze uitschieters niet aanwezig. Hoewel een exacte vergelijking met de LMR niet mogelijk is omdat daarin personen (met meerdere opnamen) meer dan 1 keer kunnen zijn opgenomen, geeft het ook voorkomen van de 1 januari en 1 juli pieken in de LMR, een indicatie dat deze data ook bij deze registratie bij voorkeur worden ingevuld, of worden opgegeven door de patiënt, als de patiënt zijn of haar werkelijke geboortedatum niet weet. De in de LMR als 'onbekend' ingevulde geboortedagen en -maanden (slechts enkele records per jaar) worden in ieder geval als '0101' (1 januari) in de registratie opgenomen.

In welke mate de registratie van een onbekende geboortedatum op persoonsniveau in de LMR op dezelfde wijze wordt gedaan als in de GBA is niet bekend. In de GBA komen echter wel geboortedata met (onbekende) geboortedag '00' voor, en soms ook met (onbekende) maand '00'. In totaal is dit bij 0,37% van de GBA-records het geval, voor het merendeel (87,3%) betreft dit in Marokko geboren personen. Het gaat hier vooral om de oudere leeftijdsgroepen; in de geboortejaren vóór 1950 is zelfs 70–100% van de in Marokko geboren personen in de GBA geregistreerd met een onbekende geboortedag. Bij opname in het ziekenhuis zullen deze personen uiteindelijk met een andere geboortedatum (veelal '0101') in de LMR worden geregistreerd. Deze patiënten zullen daarom meestal niet uniek met de GBA koppelen, of soms wel, maar dan foutief.

Door ophoging van de gekoppelde gegevens kan gecorrigeerd worden voor de verschillende koppelkansen naar geboortedag, geslacht en geboortelandgroep. Eventuele miskoppelingen kunnen echter niet worden gecorrigeerd.

Postcode

In principe moeten in de GBA woonadressen worden geregistreerd, slechts bij uitzondering kunnen briefadressen worden opgenomen. Hoewel een straatnaam-postcode zoals afgeleid kan worden uit het TPG-referentiebestand niet altijd naar een woonadres hoeft te verwijzen, kan dit wel als een grove proxy voor een woonadres beschouwd worden. Vergelijking met het TPG-referentiebestand laat zien dat de postcodes die in de GBA zijn opgenomen inderdaad voornamelijk postcodes van straatnaam adressen zijn. Op postcode-6 niveau (totaal 419 661 postcodes in de GBA) zijn er slechts 30 postcodes die naar een postbus verwijzen en 5 die naar zowel een straatnaam als een postbus verwijzen.

In de LMR komen meer postbusadressen voor dan in de GBA. Op postcode-6 niveau (totaal 270 429 postcodes in de LMR) zijn er ca. 2 100 postcodes die naar een postbus verwijzen en ca. 39 500 postcodes die naar zowel een straatnaam als een postbus verwijzen. Er zijn echter vrijwel geen LMR-records op een uitsluitend postbus-postcode gekoppeld met de GBA (3 records op postcode-4 en 1 op postcode-6 niveau over 7 jaar). De ambigue postcodes die zowel naar een postbus als een postadres verwijzen

kunnen in theorie tot miskoppelingen leiden, met name wanneer het LMR-adres een postbus betreft en deze koppelt met een straatnaam in de GBA.

5.3 Mogelijke miskoppelingen op basis van de postcode

Het probleem van miskoppelingen als gevolg van fouten in de postcode of door het koppelen van postbusadressen aan straatnaamadressen is nader onderzocht door van de records met een postcode-6 de koppelingsresultaten op postcode-6 niveau te kruisen met die op postcode-4 niveau en deze te verbijzonderen naar aard van de postcode-4. Voor deze vraagstelling is alleen uitgegaan van de LMR-records met een potentieel geldige postcode-6, bestaande uit 4 cijfers en 2 letters, en zijn koppelingen op overlidensdatum buiten beschouwing gelaten (alleen de koppelingsstappen I en II zijn meegenomen).

In tabel 11 zijn de uitkomsten voor alle typen postcodes gegeven. Bij de hierna volgende analyse wordt aangenomen dat de LMR-records met een geldige postcode-6 een representatief beeld geven van de records met alleen een postcode-4. Dit houdt ook in dat het registreren van een postcode-6 (in plaats van alleen postcode-4) door een instelling in principe onafhankelijk is van het type postcode (alleen straatnaam, alleen postbus, of beide). Onder deze aanname kan bestudeerd worden wat de effecten zijn

van het niet gebruiken van de laatste 2 posities van de postcode. Uit de rij-percentages van tabel 11 blijkt dat van de uniek gekoppelde records op postcode-6 niveau 13% meervoudig koppelt op postcode-4 niveau: deze records vallen dan af als gevolg van het beperkte discriminerende vermogen van de postcode-4 koppeling. Aan de andere kant blijkt dat van de niet gekoppelde postcode-6 records er toch 38% uniek wordt gekoppeld op postcode-4 niveau. Of, uitgaande van de kolom-percentages: 3% van de uniek-gekoppelden op postcode-4 niveau blijkt op hun postcode-6 niet te koppelen. Hiervan is het de vraag welk deel juiste koppelingen zijn (bijvoorbeeld in geval van een typefout in de laatste 2 posities van de postcode) en welk deel onjuist (bijvoorbeeld doordat een postbus met een straatnaam wordt gekoppeld of door een typefout in de viercijferige postcode). De omvang van de miskoppelingen door typefouten is niet bekend; de miskoppelingen als gevolg van 'postbusverstoring' kunnen wel nader worden onderzocht. Hiervoor is tabel 11 opgesplitst in een tabel met alleen de records met postcodes die naar zowel postbussen als straatnamen verwijzen (tabel 11a; 11% van de records) en een tabel met de postcodes die naar alleen straatnamen verwijzen (tabel 11b; 89% van de records). De tabel met de resterende records (n=7 208; 0,2% van totaal) met 'alleen postbus' postcode is niet weergegeven: van deze records koppelt er 1 op zowel postcode-6 als postcode-4 niveau en de rest koppelt op geen van beide.

Uit tabel 11a blijkt dat bij de ambigue postcodes nu 47% van de niet op postcode-6 koppelende records wel op de postcode-4 kop-

Tabel 11
Resultaten van koppeling met GBA op postcode-4 en postcode-6 niveau van LMR-records (1995-2001) met geldige 6 posities postcode

		Koppeling op postcode 4 posities							
		niet gekoppeld		uniek gekoppeld		meervoudig gekoppeld		totaal aantal records	
		abs.	(rij %)	abs.	(rij %)	abs.	(rij %)	abs.	(rij %)
Koppeling op postcode 6 posities									
Niet gekoppeld	abs. (kolom %)	166 863 (100)	(56)	113 642 (3)	(38)	19 280 (3)	(6)	299 785 (7)	(100)
Uniek gekoppeld	abs. (kolom %)	X (x)	(x)	3 542 549 (97)	(87)	544 103 (92)	(13)	4 086 652 (93)	(100)
Meervoudig gekoppeld	abs. (kolom %)	X (x)	(x)	X (x)	(x)	29 436 (5)	(100)	29 436 (1)	(100)
Totaal aantal records	abs. (kolom %)	166 863 (100)	(4)	3 676 191 (100)	(83)	592 819 (100)	(13)	4 415 873 (100)	(100)

'x': kan op logische gronden niet voorkomen.

Tabel 11a
Uitsplitsing van tabel 11 voor records waarvan de postcode (op 4 posities niveau) naar postbussen én straatnamen verwijst

		Koppeling op postcode 4 posities							
		niet gekoppeld		uniek gekoppeld		meervoudig gekoppeld		totaal aantal records	
		abs.	(rij %)	abs.	(rij %)	abs.	(rij %)	abs.	(rij %)
Koppeling op postcode 6 posities									
Niet gekoppeld	abs. (kolom %)	11 346 (100)	(50)	10 572 (2)	(47)	693 (2)	(3)	22 611 (5)	(100)
Uniek gekoppeld	abs. (kolom %)	X (x)	(x)	436 218 (98)	(94)	29 110 (88)	(6)	465 328 (95)	(100)
Meervoudig gekoppeld	abs. (kolom %)	X (x)	(x)	X (x)	(x)	3 357 (10)	(100)	3 357 (1)	(100)
Totaal aantal records	abs. (kolom %)	11 346 (100)	(2)	446 790 (100)	(91)	33 160 (100)	(7)	491 296 (100)	(100)

'x': kan op logische gronden niet voorkomen.

Tabel 11b
Uitsplitsing van tabel 11 voor records waarvan de postcode (op 4 posities niveau) naar *alleén straatnamen* verwijst

		Koppeling op postcode 4 posities							
		niet gekoppeld		uniek gekoppeld		meervoudig gekoppeld		totaal aantal records	
		abs.	(rij %)	abs.	(rij %)	abs.	(rij %)	abs.	(rij %)
Koppeling op postcode 6 posities									
Niet gekoppeld	abs. (kolom %)	148 310 (100)	(55)	103 070 (3)	(38)	18 587 (3)	(7)	226 967 (7)	(100)
Uniek gekoppeld	abs. (kolom %)	X (x)	(x)	3 106 330 (97)	(86)	514 993 (92)	(14)	3 621 323 (92)	(100)
Meervoudig gekoppeld	abs. (kolom %)	X (x)	(x)	X (x)	(x)	26 079 (5)	(100)	26 079 (1)	(100)
Totaal aantal records	abs. (kolom %)	148 310 (100)	(4)	3 209 400 (100)	(82)	559 659 (100)	(14)	3 917 369 (100)	(100)

'x': kan op logische gronden niet voorkomen.

pelt; terwijl het overeenkomstige percentage bij de 'alleen straatnaam' postcodes (tabel 11b) 38% is. Het verschil is een indicatie voor miskoppelingen als gevolg van postbusverstoring, d.w.z. 9% (is ook 10 572 uit tabel 11a gedeeld door 113 642 uit tabel 11) van de 3% records uit tabel 11 (= 0,27%) die uniek koppelen op postcode-4, maar niet koppelen op postcode-6. De postbusverstoring is dus gering, hetgeen ook blijkt uit een directe vergelijking van de percentages van tabel 11b (alleen straatnaam-adressen) met die van tabel 11. De percentages wijken nauwelijks af van elkaar, hetgeen ook de verwachting is als er weinig miskoppelingen als gevolg van postbusverstoring optreden.

Samenvattend kan gesteld worden, nog steeds onder de aanname dat de LMR-records met een geldige postcode-6 representatief zijn voor de records met alleen een postcode-4, dat door het verwaarlozen van de 2 letters van de postcode:

- 13% van de records 'verloren gaat' omdat zij meervoudig koppelen (tabel 11).
- 3% van de records als uniek koppelbaar beschouwd worden hoewel ze dat volgens het postcode-6 criterium niet zijn. Dit betreft een mengeling van juiste en niet juiste koppelingen. De verhouding hiervan is echter niet bekend, dus kan alleen geconcludeerd worden dat op grond van dit criterium minder dan 3% van de uniek op postcode-4 gekoppelden, foutieve koppelingen betreft (tabel 11). Op alle records berekend is dit iets minder (maximaal 2,6%).
- 9% van de gevallen onder (2), dat wil zeggen 9% van 3% is ca. 0,3% van de uniek gekoppelde records mogelijk foutief koppelt als gevolg van verwisselingen van postbus- en woonadressen (tabel 11a en 11).

5.4 Vergelijking van overlijdensdata in LMR en in GBA

De kwaliteit van de koppeling kan verder onderzocht worden door de waarden van variabelen die zowel in de GBA als in de LMR geregistreerd zijn, maar niet in de primaire koppeling zijn meegenomen, te vergelijken. De enige overeenkomstige variabele waarvoor dit opgaat is de overlijdensdatum: in de LMR is deze geregistreerd als patiënten tijdens een ziekenhuisopname overlijden; in de GBA is de overlijdensdatum van alle overledenen geregistreerd. In de koppelingsstappen I en II is de overlijdensdatum niet als koppelvariabele meegenomen. Door de overlijdensdata van de uniek gekoppelde LMR-records (1995–2001) na stap I en II te vergelijken met die van de gekoppelde persoon in de GBA ('RIN-Persoon' in het uit de GBA afgeleide CKP) kan de validiteit van de koppeling worden nagegaan.

Vergelijking van in LMR overledenen met overlijden volgens GBA

Van de 14 105 121 uniek gekoppelde LMR-records (na stap I en II) is er bij 295 356 records (2,1%) sprake van overlijden in het ziekenhuis. Geaggregeerd op RINPersoon niveau blijken 294 708 RINpersonen gekoppeld te zijn met 1 LMR-record, 321 RINPersonen met 2 LMR-records en 2 RINPersonen met 3 LMR-records met overlijden.

De 323 RINPersonen met 2 of meer bijbehorende LMR-records (totaal 648 'dubbel'-tellingen), dat wil zeggen de op het eerste gezicht 'meer dan eens in de LMR overleden personen', zijn nader onderzocht. Een relatief groot deel (14%) blijkt uit nuljarigen te bestaan die vrijwel allemaal op of rond de datum van geboorte zijn overleden. Bij de meerderheid (69%) van de dubbeltellingen (n=450; behorend bij 224 RINPersonen) is er sprake van gelijke instellings- en patiëntnummers bij de dubbeltellingen en/of van dezelfde datum van overlijden in zowel de LMR-records als de bijbehorende RINPersoon in de GBA. Daarom lijkt het aannemelijk dat het hier vooral gaat om meervoudige administratieve verwerking in de LMR en niet om miskoppelingen. Bij de resterende minderheid van 198 dubbeltellingen (99 RINPersonen) is de oorzaak minder duidelijk; alleen kon vastgesteld worden dat (slechts) 5 van de desbetreffende RINPersonen volgens de GBA niet waren overleden. In het algemeen kan geconcludeerd worden dat het totaal aantal 'meervoudig in de LMR overleden personen.' zeer gering is en dat het hier bovendien meestal lijkt te gaan om dubbele opnamerecords in de LMR en niet om miskoppelingen.

Van de 294 708 RINPersonen die aan 1 LMR-record (met overlijden) gekoppeld zijn, is de datum van overlijden zoals geregistreerd in de LMR vergeleken met die in de GBA. De resultaten zijn weergegeven in tabel 12.

Bij tabel 12 moet de kanttekening gemaakt worden dat voor de LMR-records van eind 2001 (met een overlijdensdatum in 2001) de opvolgperiode voor de sterfte in de GBA (tot 1 januari 2002) beperkt is geweest; op het totaal aantal gegevens zal dit echter slechts een geringe vertekening geven.

Uit tabel 12 blijkt dat in 96,8% van de records de overlijdensdatum in de LMR en GBA exact overeenstemt; als een marge van +/- 1 dag wordt toegelaten stemt zelfs 98,5% overeen. Een dergelijke marge lijkt op inhoudelijke gronden redelijk (bijvoorbeeld sterfgevallen in de nacht die in de ene registratie op de dag ervoor en in de andere registratie op de dag erna worden geregistreerd) en dit wordt ook ondersteund door de verdeling in tabel 12. Hier- van uitgaande kan dus geconcludeerd worden dat er slechts in

1,5% van de gevallen (n=4 384) een substantieel verschil (>1 dag) is gevonden in overlijdensdatum, bij 1,2% is er in de bestudeerde tijdsperiode helemaal geen overlijdensdatum in de GBA gevonden. Dit kan duiden op een (type)fout in de registratie van overlijden in de LMR of GBA, en/of op miskoppeling. Bij de niet in de GBA overleden groep (1,2%; n=3 385) hebben zich volgens de GBA bij 91% nog mutaties voorgedaan die meer dan 30 dagen na de overlijdensdatum in de LMR plaatsvonden. Dit lijkt vooral in de richting van miskoppeling te wijzen. De mutaties betroffen overigens vooral 'emigratie' (ongeveer de helft van de gevallen). Een verklaring hiervoor ligt niet zo direct voor de hand.

Tabel 12
Verschil datum overlijden LMR en datum overlijden GBA bij uniek gekoppelde records (1995–2001), na koppelingsstap I en II, waarbij overlijden is geregistreerd in de LMR

	Abs.	%
Niet overleden in GBA	3 385	1,15
– 366 dagen en meer	99	0,03
– 91 t/m –365 dagen	32	0,01
– 61 t/m –90 dagen	20	0,01
– 31 t/m –60 dagen	39	0,01
– 15 t/m –30 dagen	47	0,02
– 8 t/m –14 dagen	63	0,02
– 7 dagen	14	0,00
– 6 dagen	17	0,01
– 5 dagen	30	0,01
– 4 dagen	76	0,03
– 3 dagen	128	0,04
– 2 dagen	257	0,09
– 1 dag	1 788	0,61
geen verschil	285 183	96,77
+ 1 dag	3 353	1,14
+ 2 dagen	63	0,02
+ 3 dagen	30	0,01
+ 4 dagen	8	0,00
+ 5 dagen	1	0,00
+ 6 dagen	2	0,00
+ 7 dagen	4	0,00
+ 8 t/m 14 dagen	21	0,01
+ 15 t/m 30 dagen	28	0,01
+ 31 t/m 60 dagen	12	0,00
+ 61 t/m 90 dagen	3	0,00
+ 91 t/m 365 dagen	5	0,00
Totaal	294 708	100,00

Uit tabel 12 blijkt dat de extreme waarden zich vooral voor doen bij een negatief verschil. In deze gevallen is iemand volgens de LMR al lang overleden, terwijl deze persoon volgens de GBA nog in leven is. Andersom komt aanzienlijk minder voor, hetgeen te verwachten is, aangezien het eerdere overlijden in de GBA zich

alleen binnen de beperkte periode van de opnamedatum minus 30 dagen tot de ontslagdatum (overlijdensdatum) in de LMR kan voordoen. De LMR-records zijn immers met de GBA gekoppeld met een marge van 30 dagen rond de opnamedatum.

Verder is nog onderzocht in hoeveel gevallen de datum van overlijden in de GBA vóór de datum van opname in de LMR ligt. Dit blijkt bij slechts 72 personen het geval te zijn. Meestal gaat het hier om een verschil van slechts 1 dag en is de opnamedatum gelijk ook de overlijdensdatum. De overige verschillen wijzen ook niet direct op miskoppeling, maar meer in de richting van intoetsfouten bij de overlijdensdatum.

Vergelijking van niet in LMR overledenen met overlijden in de GBA

Van de uniek gekoppelde LMR-records waarbij géén overlijden is geregistreerd tijdens de opname (n=13 809 765) is nagegaan of er ook geen overlijden is geregistreerd in de GBA. Daarbij is in de GBA alleen gekeken naar overlijdens op of voor de ontslagdatum omdat alleen deze overlijdens eventueel kunnen duiden op miskoppeling. Immers het omgekeerde, overlijden in de GBA na de ontslagdatum, is heel goed mogelijk: bijna 60% van de overlijdensgevallen vindt niet in het ziekenhuis plaats. De resultaten van de vergelijking staan vermeld in tabel 13.

Uit tabel 13 blijkt er van de 13 809 765 LMR-records, behorende bij 6 221 187 RINPersonen, slechts bij 280 records volgens de GBA een overlijden heeft plaatsgevonden vóór de ontslagdatum in de LMR, een verwaarloosbaar klein aantal. Uitgaande van een toelaatbaar verschil in overlijdensdatum van +/- 1 dag blijven er hiervan nog maar 188 records over waarbij mogelijk sprake is van miskoppeling. Bij een grotere groep records (n=5 530, maar relatief nog steeds heel klein (0,04%), is een overlijden in de GBA geregistreerd op de ontslagdatum: dit hoeft echter niet noodzakelijkerwijs te wijzen op miskoppeling omdat het mogelijk is dat een patiënt elders komt te overlijden op de dag dat hij of zij is ontslagen uit het ziekenhuis.

Conclusies

Uit dit vergelijkend onderzoek naar de consistentie van de overlijdensgevallen zoals geregistreerd in de LMR en in de GBA kan geconcludeerd worden dat de kwaliteit van de koppeling op grond van dit criterium goed is. Vrijwel alle, volgens de LMR in het ziekenhuis overleden, personen (98,5%) hebben een overlijdensdatum in de GBA die gelijk is aan, of maximaal 1 dag verschilt met, die in de LMR. Bij de overige 1,5% kan geen duidelijk onderscheid gemaakt worden tussen miskoppelingen, fouten in overlijdens- of emigratiegegevens, of typefouten. Verder zijn van de niet in de LMR overleden personen er meer dan 99,99% ook niet (voor de ontslagdatum) overleden volgens de GBA-gegevens.

Tabel 13
Overlijdensgegevens uit de GBA behorende bij uniek gekoppelde LMR records (1995–2001), na koppelingsstap I en II, waarbij in de LMR geen overlijden is geregistreerd

	Abs.	%
Totaal aantal uniek gekoppelde LMR records	13 809 765	100,00
w.v.		
GBA overlijden op LMR ontslagdatum	5 530	0,04
GBA overlijden vóór LMR ontslagdatum	280	0,00
w.v.		
GBA overlijden op ontslagdatum minus 1 dag	92	0,00
GBA overlijden in periode van opnamedatum tot ontslagdatum minus 1 dag	71	0,00
GBA overlijden vóór opnamedatum	117	0,00

5.5 Onwaarschijnlijke mutaties in de GBA-gegevens tijdens LMR-opnameperiode

De kwaliteit van de koppeling kan ook onderzocht worden door bij de uniek gekoppelde records na te gaan of er voor of tijdens de LMR-opname onwaarschijnlijke mutaties – anders dan sterfte – in de GBA-gegevens voorkomen. Adreswijzigingen zijn hier niet onderzocht, omdat deze zich in principe wel (administratief) kunnen voordoen tijdens of direct voor of na een opname. De mutaties die zijn onderzocht betreffen emigratie en wijziging van burgerlijke staat. Bij dit onderzoek is uitgegaan van alle uniek gekoppelde records na stap III (n=14 121 787).

Emigratie vóór of tijdens de opname

Van de 14 121 787 uniek gekoppelde LMR-records blijkt er bij 329 sprake te zijn van emigratie vóór de opname en bij 117 van emigratie tijdens de opnameperiode. In totaal betreft dit dus 446 records, dat wil zeggen slechts 0,003%.

Wijziging van burgerlijke staat tijdens de opname

Van alle uniek gekoppelde records is er bij 3 059 volgens de GBA sprake van wijziging van burgerlijke staat tijdens de opnameperiode, dat wil zeggen vanaf de opnamedatum t/m de ontslagdatum. Deze groep bestaat echter voor het grootste deel (60%; n=1 859) uit verweduwden: personen die in een ziekenhuis verbleven in de periode dat hun huwelijkspartner overleed. Deze situatie is geheel plausibel. Bij 18% (n=542) is er sprake van echtscheiding. Ook dit is niet onwaarschijnlijk omdat bij de formele administratieve verwerking van een echtscheiding de huwelijkspartners niet fysiek aanwezig zijn. De meest onwaarschijnlijke mutatie – hoewel ook dit strikt genomen niet onmogelijk is – betreft een groep van 130 personen die tijdens de opnameperiode in het huwelijk treden. Op de totale groep van 14 121 787 uniek gekoppelde records is dit aantal echter verwaarloosbaar klein.

Conclusies

Geconcludeerd kan worden dat bij de uniek gekoppelde records op grond van de GBA-gegevens vrijwel geen onwaarschijnlijke mutaties – anders dan sterfte – hebben plaatsgevonden tijdens of voor de opnameperiode. Ook dit is een aanwijzing voor een goede kwaliteit van de koppeling.

5.6 Overlappende opnameperioden bij dezelfde persoon

Een ander aspect dat inzicht kan geven in de kwaliteit van het koppelingsproces is de mate van vóórkomen van overlappende opnameperioden bij dezelfde persoon. Overlappende opnamen kunnen namelijk op miskoppeling wijzen, maar dit dient eerst nader genuanceerd te worden. Als gevolg van de wijze waarop

ziekenhuizen opnamen administratief vastleggen, kan het bijvoorbeeld voorkomen dat de dag van opname in het bestemmingsziekenhuis overeenkomt met de dag van ontslag in het ziekenhuis van herkomst. In de praktijk hoeven overlappende opnamen daarom niet zonder meer te duiden op administratieve onjuistheden of miskoppeling. Vooral dagopname in een instelling, gevolgd door een klinische opname in dezelfde instelling op dezelfde dag, komt regelmatig voor. Ook kan er tijdens een klinische opname een dagopname zijn in een andere instelling, bijvoorbeeld voor een dotterbehandeling. Daarom is bij dit onderzoek de volgende definitie gehanteerd voor overlappende opnamen: 'opnamen gekoppeld aan dezelfde persoon waarbij de opnameperiode van de ene opname meer dan 1 dag overlapt met de opnameperiode van de andere opname'.

Uitgaande van alle uniek gekoppelde records (na stap III) geeft toepassing van voornoemde definitie als resultaat dat 0,12% van de LMR-records overlap vertonen (zie tabel 14).

Uit tabel 14 blijkt dat er een groep is van 8 427 personen waarbij sprake is van 2 of meer overlappende opnamen. In totaal gaat het om 17 024 opnamen, waarvan het merendeel (97,1%) betrekking heeft op 2 elkaar overlappende opnamen. De overige 2,9% betreft 3 of meer opnamen.

Deze 17 024 overlappende opnamerecords (0,12% van het totaal aantal opnamen) worden hieronder nader gekarakteriseerd.

Overlappende opnamen op instellingsniveau

Als gekeken wordt naar de verdeling van overlappende opnamen over de afzonderlijke instellingen dan blijken er verschillen te zijn in de fractie overlappende opnamen ten opzichte van het totaal aantal opnamen per instelling. Bij de uniek-gekoppelden varieert deze fractie van 0,02 tot 0,71%. Voor alle records waarbij sprake is van 2 overlappende opnamen (n=16 532; 97,1% van het totaal), is nader onderzocht of er sprake is van overlap binnen dezelfde instelling of van overlap tussen twee verschillende instellingen. Hieruit blijkt dat van deze overlappende opnamen er bij 5 042 (30,5%) overlap is binnen dezelfde instelling. Het betreft hier dus overlappende opnameperioden van 2 of meer dagen. Bij deze overlappende opnamen binnen dezelfde instelling is nagegaan in hoeverre ook het patiëntnummer overeenstemt. Bij 75% (n=3 780) van de opnamen is sprake van verschillende patiëntnummers binnen dezelfde instelling, wat mogelijk zou kunnen duiden op miskoppeling. Hierbij wordt er dan vanuit gegaan dat de desbetreffende ziekenhuizen bij heropnamen van dezelfde patiënt hetzelfde patiëntnummer toekennen; dit hoeft evenwel niet altijd het geval te zijn. Bij de overige 25% komt het patiëntnummer overeen, hetgeen, onder dezelfde aanname als hiervoor, een indicatie is dat het hier wel om dezelfde personen gaat. De overlappende opnamen zijn dan ontstaan door fouten in de opname/ontslagdatum of door onjuiste toepassing van de LMR-regels.

Van de 16 532 overlappende records die betrekking hebben op 2 opnamen is tevens nagegaan of de typering van de bestemming

Tabel 14
Overlappende opnamen bij uniek gekoppelde LMR records (1995–2001)

Totaal aantal records	14 121 787	(bij 6 335 793 RINPersonen)
Overlappende opname records	17 024	(bij 8 427 RINPersonen)
w.v.		
2 overlappende opname records	16 532	
3 of meer overlappende opname-records	492	
Fractie overlappende records	0,12%	

en herkomst van de patiënt, zoals geregistreerd in de records, in overeenstemming is met elkaar. Ook hierbij is onderscheid gemaakt tussen overlap met een andere instelling (n=11 490) en overlap binnen dezelfde instelling (n=5 042).

Overlap met een andere instelling

De 11 490 overlappende opnamen tussen 2 verschillende instellingen zijn op persoonsniveau paarsgewijs vergeleken op de bestemmings- en herkomstvariabele, waarbij is uitgegaan van de chronologische volgorde. Dit houdt in dat van de eerste opname (n=5 745) is gekeken naar de bestemming en de bestemmingsinstelling en dat bij de tweede opname (n=5 745) gekeken is naar de herkomst en de herkomstinstelling. Bij de bestemming is ook onderscheid gemaakt tussen overlappende opnamen waarbij de tweede opname binnen de opname periode van de eerste valt (n=3 057) en overlappende opnamen waarbij de tweede opname langer doorloopt dan de eerste (n=2 688); de resultaten hiervan staan vermeld in tabel 15.

Als de overlappende opnamen daadwerkelijk bij dezelfde persoon horen, zou men op het eerste gezicht verwachten dat bij de eerste opname als bestemming 'andere instelling' hoort te zijn ingevuld. Uit tabel 15 blijkt dat dit bij 33% ook inderdaad het geval is. Daarvan is nagegaan of ook inderdaad het juiste instellingsnummer bij de bestemming is opgegeven, dat wil zeggen hetzelfde instellingsnummer als de instelling waarbij de tweede opname plaats vond. Omdat het opgeven van een instellingsnummer een facultatieve variabele is dat niet door alle ziekenhuizen wordt ingevuld kon dit alleen voor een deel van de records worden nagegaan (totaal 977 van de 1 888 records met bestemming 'andere instelling'). Bij 86% (n=843) van deze records was inderdaad het juiste instellingsnummer ingevuld. Als dit percentage ook toegepast wordt op de opnamen met een onbekend instellingsnummer en gepercentreerd wordt op alle bestemmingen, dan is het bij 28% (86% x 1 888 / 5 745) van de opnamen zeer aannemelijk dat de overlappende opnamen inderdaad bij dezelfde persoon horen en dat er dus geen sprake is van miskoppeling.

Dat betekent verder niet dat de rest van de overlappende opnamen noodzakelijkerwijs wel het gevolg van miskoppeling zouden zijn. Uit tabel 15 valt bijvoorbeeld op dat de bestemming 'eigen woonomgeving' vaker wordt opgegeven als de tweede opname binnen de eerste opnameperiode valt, dus voor of op de ontslagdatum van de eerste opname eindigt. Dit percentage is aanzienlijk lager bij tweede opnamen die nog voortduren als de eerste al afgelopen is. Dit kan er op wijzen dat in het eerste geval bij ten minste een deel van de opnamen de bestemming pas wordt toegekend nadat de persoon weer is teruggekeerd naar de herkomstinstelling. In dat geval is er ook geen sprake van miskoppeling.

Vervolgens is gekeken naar de variabele herkomst bij de tweede opname. Hier zou men ook verwachten dat, indien de overlappende opnamen bij dezelfde persoon horen, altijd 'andere instelling' is ingevuld. Uit tabel 16 blijkt dat dit bij 45% ook het geval is.

Tabel 16
Herkomst van tweede opname bij overlap met eerste opname in andere instelling, bij 5 745 paren van records (1995–2001)

	Abs.	%	
Herkomst			
Eigen woonomgeving	2 398	41,7	
Bejaardenoord	2	0,0	
Andere instelling	2 610	45,4	
w.v. % 'juist' nr. ¹⁾			91%
w.v. % 'onjuist' nr. ¹⁾			9%
In ziekenhuis geboren	735	12,8	
Totaal	5 745	100,0	

¹⁾ Percentage van de records met ingevuld instellingsnummer waarbij het instellingsnummer van de herkomst overeenkomt ('juist' nr.), resp. niet overeenkomt ('onjuist' nr.) met het instellingsnummer van de 1e opname.

Van deze 45% is weer nagegaan wat is ingevuld bij de instelling van herkomst. Hieruit blijkt dat bij 91% (947 van de in totaal 1 040 records waarbij instellingsnummer was ingevuld) de herkomstinstelling overeenkomt met de instelling van de eerste opname. Van het totaal aantal record-paren wijst derhalve in dit geval 41% in ieder geval niet op miskoppeling. Ook hier hoeft er bij de overige records niet altijd sprake te zijn van miskoppeling, b.v. als de opnamedatum van de tweede opname op of vlak na de opnamedatum van de eerste opname ligt en men bij de tweede opname dezelfde herkomst als van de eerste opname heeft ingevuld.

Overlap met dezelfde instelling

Voor overlappende opnamen die betrekking hebben op dezelfde instelling is een soortgelijke exercitie uitgevoerd; de resultaten daarvan staan vermeld in Bijlage 3.

Overlappende opnamen naar diagnose

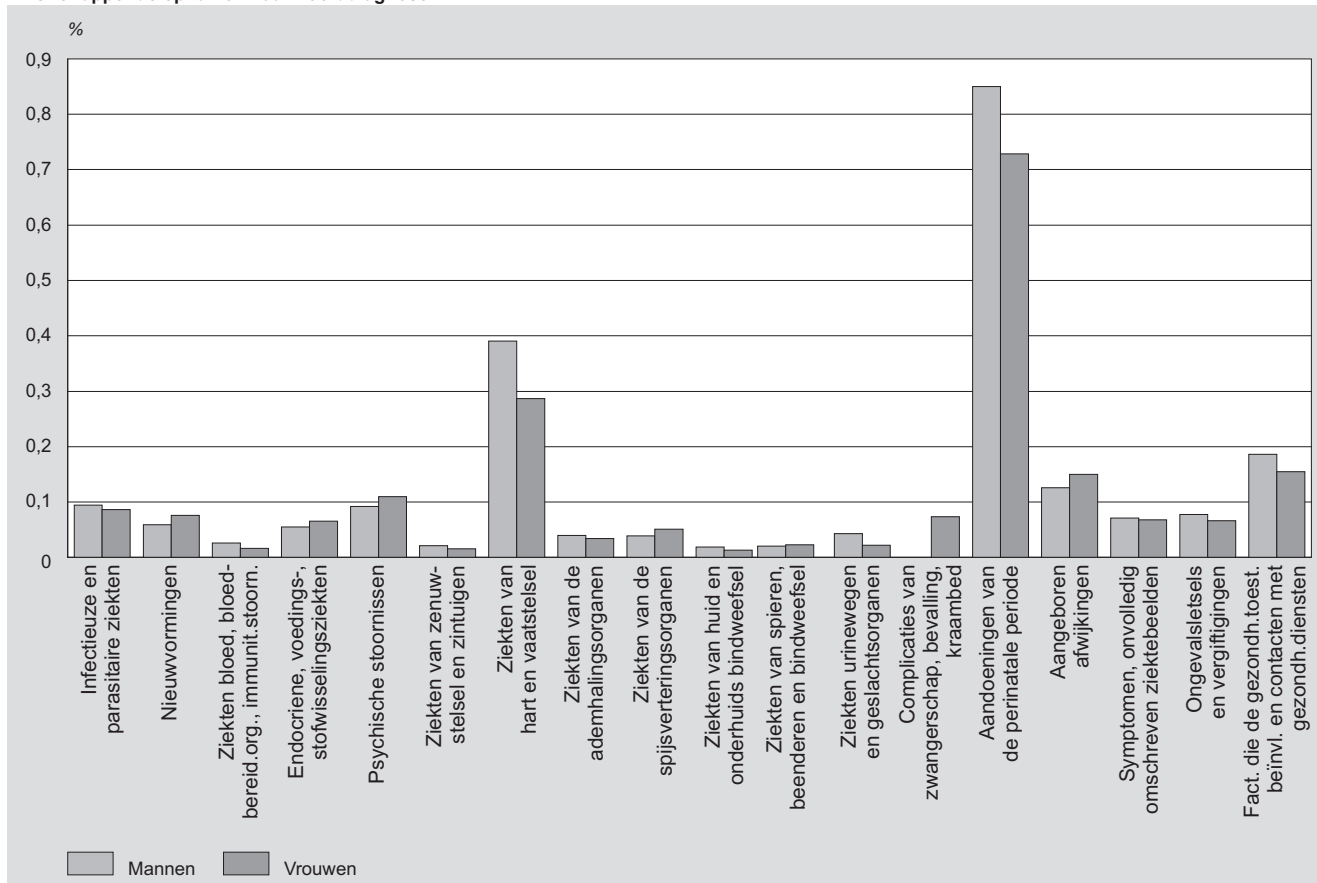
Om meer zicht te krijgen in de aard van de overlappende opnamen zijn deze ook onderzocht naar diagnosegroep.

Tabel 15
Bestemming na eerste opname bij overlap met tweede opname in andere instelling, bij 5 745 paren van records (1995–2001)

Bestemming	Totaal		2e opname valt binnen 1e opn. periode		2e opname eindigt na 1e opname		
	abs.	%	abs.	%	abs.	%	
Eigen woonomgeving	3 761	65,5	2 666	87,2	1 095	40,7	
Bejaardenoord	13	0,2	8	0,3	5	0,2	
Andere instelling	1 888	32,9	311	10,2	1 577	58,7	
w.v. % 'juist' nr. ¹⁾							86%
w.v. % 'onjuist' nr. ¹⁾							14%
In instelling overleden	79	1,4	68	2,3	11	0,4	
Tegen advies weggegaan	4	0,1	4	0,1	–	–	
Totaal	5 745	100,0	3 057	100,0	2 688	100,0	

¹⁾ Percentage van de records met ingevuld instellingsnummer waarbij het instellingsnummer van de bestemming overeenkomt ('juist' nr.), resp. niet overeenkomt ('onjuist' nr.) met het instellingsnummer van de 2e opname.

2. Overlappende opnamen naar hoofddiagnose



Als de hoofddiagnosen van alle overlappende opnamen (17 024 records) vergeleken worden met hoofddiagnosen van alle uniek gekoppelde opnamen (14 121 787 records) dan blijkt dat er bij de diagnosegroep 'ziekten van hart en vaatstelsel' en bij de groep 'aandoeningen van de perinatale periode' sprake is van een sterk verhoogd percentage overlappende opnamen (zie figuur 2), zowel bij mannen als bij vrouwen, ten opzichte van het te verwachten gemiddeld percentage van 0,12% (0,14% bij mannen en 0,10% bij vrouwen).

Bij ziekten van hart en vaatstelsel is het percentage overlappende opnamen 0,35% en bij aandoeningen van de perinatale periode is dit zelfs 0,79%. Dit duidt er op dat het verschijnsel overlappende opnamen samen lijkt te hangen met de aard van de diagnose. Bij deze groepen van aandoeningen is het ook aannemelijk dat er relatief veel overdrachten plaatsvinden, ook naar andere instellingen.

Als overlappende opnamen nader uitgesplitst worden naar de opnamereden blijkt dat bij aandoeningen van de perinatale periode er in verhouding vaak sprake is van observatie (ca. 32%) t.o.v. 12% voor alle diagnoses. Ook wordt bij aandoeningen van de perinatale periode relatief vaak (9%) gastverblijf als opname reden gegeven. Wellicht wordt hiermee aangegeven dat na een opname in verband met geboorte, een zuigeling soms administratief nogmaals wordt opgenomen voor 'observatie', of 'gastverblijf' bij de zieke moeder.

Van de opnamerecords die betrekking hebben op 2 elkaar overlappende opnamen (n=16 532) is verder nog nagegaan of er verschil is tussen overlap binnen dezelfde instelling en overlap met andere instellingen wat betreft de hoofddiagnose van de opname.

Uit tabel 17 blijkt dat bij de overlappende opnamen zowel binnen dezelfde instelling als met een andere instelling, er vaker sprake

van 'aandoeningen van de perinatale periode' en 'ziekten van hart en vaatstelsel' in vergelijking met de aandelen van deze diagnoses bij het totaal van uniek gekoppelde records. Bij de overlappen binnen dezelfde instelling is vooral een sterke oververtegenwoordiging van de aandoeningen van de perinatale periode te zien, terwijl bij de overlappen met een andere instelling verhoudingsgewijs vooral de ziekten van hart en vaatstelsel eruit springen. Bij deze laatste categorie van aandoeningen is er blijkbaar relatief vaak sprake van (al of niet tijdelijke) overplaatsing naar een andere instelling, b.v. een voor een bepaalde cardiovasculaire ingreep gespecialiseerde instelling. Overigens komt ook bij de aandoeningen van de perinatale periode overlappen met een andere instelling vaker voor dan bij het totaal van de unieke gekoppelde records. De hoofddiagnosegroep 'Factoren die de gezondheidstoestand beïnvloeden en contacten met gezondheidsdiensten' is verder alleen binnen dezelfde instelling oververtegenwoordigd.

Er van uitgaande dat de extra vertegenwoordiging van de ziekten van hart en vaatstelsel (+24%) en van de perinatale periode (+15%) bij de overlappende opnamen in ieder geval niet wijzen op miskoppeling, dan zou dit in totaal bij 39% van de overlappende opnamen het geval zijn. Maximaal 61% van de overlappende opnamen zou dan wel kunnen wijzen op miskoppelingen; op het totaal aantal uniek gekoppelde opnamen is dit 0,07%.

Concluderend kan gesteld worden dat de ruime oververtegenwoordiging van de hoofddiagnosen 'aandoeningen van de perinatale periode' en 'ziekten van hart en vaatstelsel' bij de overlappende opnamen er op wijst dat hier voor een aanzienlijk deel (ca. 40% of meer) waarschijnlijk niet sprake is van miskoppelingen. Als overlap zou samenhangen met foutieve koppeling dan zou er immers verwacht mogen worden dat dit onafhankelijk is van de diagnose. Bij de genoemde diagnoses is het ook aannemelijk dat hierbij relatief veel overdrachten plaatsvinden, die voor een deel

Tabel 17

Procentuele verdeling hoofddiagnosen van overlappende opnamen, binnen zelfde of met andere instelling, en van alle uniek gekoppelde opnamen, 1995–2001

	Overlappende opnamen			Alle opnamen
	binnen zelfde instelling	met andere instelling	totaal	
	%			
Hoofddiagnose				
Infectieuze en parasitaire ziekten	0,8	0,7	0,7	1,0
Nieuwvormingen	4,1	5,6	5,2	8,8
Ziekten van bloed, bloedbereid.org., immunit.stoorn.	0,4	0,1	0,2	1,0
Endocriene, voedings-, stofwisselingsziekten	1,2	0,7	0,8	1,6
Psychische stoornissen	1,0	0,8	0,9	1,0
Ziekten van zenuwstelsel en zintuigen	1,4	1,5	1,4	9,4
Ziekten van hart en vaatstelsel	7,4	49,0	36,3	12,3
Ziekten van de ademhalingsorganen	2,4	2,3	2,3	7,4
Ziekten van de spijsverteringsorganen	3,7	2,5	2,8	7,4
Ziekten van huid en onderhuids bindweefsel	0,3	0,1	0,2	1,4
Ziekten van spieren, beenderen en bindweefsel	3,5	1,2	1,9	10,2
Ziekten urinewegen en geslachtsorganen	2,4	1,2	1,5	6,0
Complicaties van zwangerschap, bevalling, kraambed	5,3	3,3	3,9	6,2
Aandoeningen van de perinatale periode	29,6	11,6	17,1	2,5
Aangeboren afwijkingen	0,8	1,3	1,2	1,0
Symptomen, onvolledig omschreven ziektebeelden	4,0	3,1	3,4	5,7
Ongevalsletsels en vergiftigingen	3,7	3,2	3,4	5,5
Fact. die de gezondh.toest. beïnv. en contacten met gezondh.diensten	28,0	11,8	16,7	11,6
Totaal (abs.=100%)	5 042	11 490	16 532	14 121 787

dus als nieuwe opname bij dezelfde persoon geregistreerd lijken te worden.

Conclusies overlappende opnamen

Samengevat kan worden geconcludeerd dat het vóórkomen van overlappende opnamen bij dezelfde persoon beperkt is tot 0,12% van het totaal aantal uniek-gekoppelde records. Nader onderzoek heeft echter uitgewezen dat een deel van deze overlappende opnamen niet wijst in de richting van miskoppeling. Dit deel is niet precies te kwantificeren, maar zou naar schatting minimaal 25–40% kunnen bedragen, afhankelijk van wat men als meetpunt neemt (gelijke patiëntnummers bij overlappende opnamen in dezelfde instelling, congruente bestemming en herkomst variabelen bij overlappende opnamen bij verschillende instellingen, of oververtegenwoordiging van bepaalde aandoeningen). Het percentage opnamen dat wel indicatief zou kunnen zijn voor miskoppelingen daalt dan, berekend op alle opnamen, ruim onder de 0,1%.

5.7 Kenmerken van de niet gekoppelde LMR-records

In paragraaf 5.1 en 5.2 zijn al enkele redenen aangegeven voor het niet-koppelen van sommige LMR-records met de GBA: onbekende geboortedata in de LMR en in de GBA en het niet voorkomen van sommige LMR-postcodes in de GBA (0,4% van alle LMR-records).

In totaal kunnen deze oorzaken leiden tot niet-koppelingen bij naar schatting minder dan een procent van het totaal aantal LMR-records. Het totaal aantal niet gekoppelde records bedraagt echter gemiddeld 3,6%. Een belangrijk deel hiervan zal vermoedelijk het gevolg zijn van administratieve fouten in beide registraties, waarbij bijvoorbeeld wel een geldig adres wordt opgegeven maar niet het actuele woonadres van de betreffende persoon. Van belang is of het niet-koppelen een redelijk 'random' gebeuren is of dat de niet gekoppelde LMR-records een specifieke groep betreffen. De eventuele selectiviteit van de niet gekoppelde records en de variatie daarin is nader onderzocht door achtergrondkenmerken zoals afgeleid uit de LMR te vergelijken en mogelijke fluctuaties in de tijd na te gaan. Als referentie is daarbij de groep overige LMR-records gebruikt, d.w.z. de uniek- én meervoudig-gekoppelden. In latere analyses zullen weliswaar uitsluitend de uniek-gekoppelden gebruikt worden maar deze zullen worden gewogen voor de koppelskans, waardoor in feite de meervoudig gekoppelde records verdisconteerd worden in het resultaat.

De volgende kenmerken zijn onderzocht: leeftijd, geslacht, regio volgens postcode-2 van de patiënt en instelling van opname, type zorgverlening, verpleegduur en hoofddiagnose.

Leeftijd

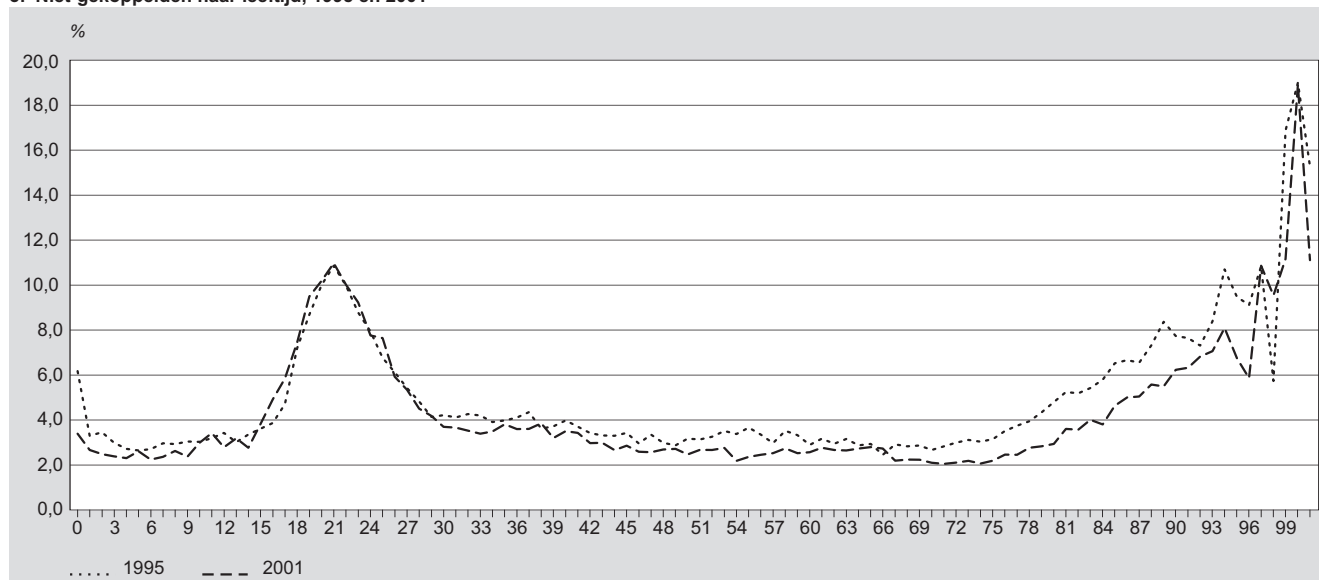
In tabel 18 en figuur 3 staan respectievelijk de gemiddelde leeftijd en de leeftijdsverdeling voor de niet gekoppelde records. Hieruit blijkt dat in alle jaren de niet-gekoppelden gemiddeld 2–3 jaar

Tabel 18

Gemiddelde leeftijd van LMR-patiënten naar koppelstatus en registratiejaar (1995–2001)

	Niet gekoppelde records		Overige records		Totaal LMR	
	gem. leeftijd (jr)	aantal	gem. leeftijd (jr)	aantal	gem. leeftijd (jr)	aantal
1995	42,7	90 186	44,6	2 099 654	44,6	2 189 840
1996	43,0	86 857	45,1	2 156 796	45,0	2 243 653
1997	43,5	84 880	45,6	2 204 298	45,5	2 289 178
1998	43,0	84 698	45,6	2 225 653	45,5	2 310 351
1999	42,0	81 340	46,2	2 249 657	46,0	2 330 997
2000	41,9	79 651	46,5	2 253 183	46,3	2 332 834
2001	43,7	79 494	46,7	2 339 265	46,6	2 418 759
Totaal	42,8	587 106	45,8	15 528 506	45,7	16 115 612

3. Niet-gekoppelden naar leeftijd, 1995 en 2001



jonger zijn dan de rest van de LMR-records, waarbij het leeftijdsverschil in de tijd iets toeneemt. Uit figuur 3 blijkt dat de niet-gekoppelden relatief vaak voorkomen bij 0-jarigen, 18–28-jarigen en bij de (zeer) hoge leeftijden.

Geslacht

Bij de niet-gekoppelde records is er voor alle jaren sprake van een hoger percentage mannen in vergelijking met de overige LMR-records. Bij de niet-gekoppelde records is dit gemiddeld 48,2% en bij de overige records 45,6%. Een verschil van 2,6 procentpunt dus over alle jaren. Het verschil neemt van 1995 tot 1997 toe van 1,4 tot 3,0% en blijft in de volgende jaren rond de 3% schommelen.

Regio

Om een indruk te krijgen van eventuele regionale verschillen in het percentage niet-gekoppelden is een onderverdeling gemaakt op basis van de eerste 2 cijfers van de postcode. De gebieden met de hoogste percentages niet-gekoppelden betreffen voornamelijk de regio's Den Haag (7,6%), Amsterdam (6,5%), Utrecht (6,5%) en, in iets mindere mate, Rotterdam (5,3%). Het gemiddelde percentage over alle regio's is 3,6%. Het gaat hier dus vooral om een concentratie in de grote steden in het westen des lands. Als gekeken wordt naar de vestigingsplaats van de instel-

lingen met een relatief hoog percentage niet-gekoppelde records dan wordt er ook een clustering in Den Haag, Utrecht en Amsterdam waargenomen.

Type zorgverlening en verpleegduur

Het is een bekend gegeven dat het aandeel dagverplegingsopnamen in het totaal van opnamen is toegenomen in de afgelopen jaren. Uit tabel 19 blijkt dat deze trend in de tijd zowel zichtbaar is bij de niet-gekoppelde records als bij de overige LMR-records. Het aandeel van de dagverpleging is bij de niet-gekoppelden echter wel gemiddeld 9,5 procentpunt lager dan bij de overigen. Bij de afzonderlijke jaren varieert dit verschil van 7,9–10,2%, maar van een trend in de tijd is hier geen sprake.

Doordat het aandeel dagverpleging bij de niet-gekoppelde records lager is en dus het aandeel klinische opnamen hoger, is het te verwachten dat de gemiddelde verpleegduren ook langer is bij deze groep. Bij de niet-gekoppelde records blijkt deze gemiddeld 10,0 dagen te zijn en bij de overige records 6,1 dagen. Wanneer alleen gekeken wordt naar de klinische opnamen dan is de gemiddelde verpleegduren voor de niet-gekoppelde records en de overige records resp. 13,0 en 8,8 dagen. In beide gevallen is er sprake van een dalende trend in de periode 1995–2001.

Geconcludeerd kan worden dat bij de niet-gekoppelde records verhoudingsgewijs vaker sprake is van klinische opnamen en een langere verpleegduren.

Tabel 19
Percentage dagverplegingsopnamen in de LMR naar koppelstatus en registratiejaar (1995–2001)

	Niet gekoppelde records		Overige records		Totaal LMR	
	% dagverpl.	abs.=100%	% dagverpl.	abs.=100%	% dagverpl.	abs.=100%
1995	20,6	90 186	28,9	2 099 654	28,6	2 189 840
1996	22,6	86 857	30,5	2 156 796	30,2	2 243 653
1997	23,0	84 880	32,3	2 204 298	32,0	2 289 178
1998	24,5	84 698	34,3	2 225 653	33,9	2 310 351
1999	25,9	81 340	35,9	2 249 657	35,5	2 330 997
2000	27,6	79 651	37,9	2 253 183	37,5	2 332 834
2001	31,1	79 494	40,5	2 339 265	40,2	2 418 759
Totaal	24,9	587 106	34,5	15 528 506	34,1	16 115 612

Tabel 20a
Verdeling van hoofddiagnosen in de LMR naar koppelstatus en registratiejaar, 1995–2001

	1995		1996		1997	
	niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld
	%					
Hoofddiagnosen						
Infectieuze en parasitaire ziekten	1,7	1,0	1,7	1,0	1,7	1,0
Nieuwvormingen	6,0	8,2	6,5	8,2	7,0	8,4
Ziekten bloed, bloedbereid.org., immuniteit.stoorn.	1,0	1,0	1,1	1,0	1,2	1,0
Endocriene, voedings-, stofwisselingsziekten	2,2	1,6	2,2	1,7	2,1	1,6
Psychische stoornissen	3,1	1,0	2,2	1,0	2,2	1,0
Ziekten van zenuwstelsel en zintuigen	6,9	9,1	6,5	9,2	6,6	9,3
Ziekten van hart en vaatstelsel	11,0	12,5	11,3	12,4	11,4	12,3
Ziekten van de ademhalingsorganen	7,7	8,2	7,9	8,1	7,4	7,7
Ziekten van de spijsverteringsorganen	7,9	7,6	7,9	7,5	7,9	7,5
Ziekten van huid en onderhuids bindweefsel	1,4	1,3	1,5	1,3	1,6	1,3
Ziekten van spieren, beenderen en bindweefsel	8,2	11,0	7,8	10,8	7,4	10,5
Ziekten urinewegen en geslachtsorganen	5,6	6,7	5,4	6,4	5,4	6,2
Complicaties van zwangerschap, bevalling, kraambed	6,2	6,1	6,2	6,2	6,0	6,2
Aandoeningen van de perinatale periode	3,9	2,5	3,8	2,6	4,0	2,5
Aangeboren afwijkingen	1,2	1,1	1,0	1,1	1,1	1,1
Symptomen, onvolledig omschreven ziektebeelden	5,5	5,0	5,8	5,0	6,4	5,6
Ongevalsextsels en vergiftigingen	10,1	5,7	10,1	5,7	10,0	5,7
Fact. die de gezondh.toest. beïnv. en contacten met gezondh.diensten	10,3	10,4	11,1	10,8	10,7	11,1
Totaal (abs.=100%)	90 186	2 099 654	86 857	2 156 796	84 880	2 204 298

Tabel 20b
Aandeel van niet-gekoppelde en overige LMR records naar hoofddiagnose en registratiejaar, 1995–2001

	1995		1996		1997	
	niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld
	%					
Hoofddiagnosen						
Infectieuze en parasitaire ziekten	6,9	93,1	6,2	93,8	6,3	93,7
Nieuwvormingen	3,1	96,9	3,1	96,9	3,1	96,9
Ziekten bloed, bloedbereid.org., immuniteit.stoorn.	4,2	95,8	4,2	95,8	4,2	95,8
Endocriene, voedings-, stofwisselingsziekten	5,6	94,4	5,0	95,0	4,7	95,3
Psychische stoornissen	11,7	88,3	8,0	92,0	7,5	92,5
Ziekten van zenuwstelsel en zintuigen	3,1	96,9	2,8	97,2	2,7	97,3
Ziekten van hart en vaatstelsel	3,7	96,3	3,5	96,5	3,5	96,5
Ziekten van de ademhalingsorganen	3,9	96,1	3,8	96,2	3,6	96,4
Ziekten van de spijsverteringsorganen	4,2	95,8	4,0	96,0	3,9	96,1
Ziekten van huid en onderhuids bindweefsel	4,5	95,5	4,4	95,6	4,4	95,6
Ziekten van spieren, beenderen en bindweefsel	3,1	96,9	2,8	97,2	2,6	97,4
Ziekten urinewegen en geslachtsorganen	3,5	96,5	3,3	96,7	3,2	96,8
Complicaties van zwangerschap, bevalling, kraambed	4,2	95,8	3,9	96,1	3,6	96,4
Aandoeningen van de perinatale periode	6,2	93,8	5,7	94,3	5,7	94,3
Aangeboren afwijkingen	4,6	95,4	3,7	96,3	3,8	96,2
Symptomen, onvolledig omschreven ziektebeelden	4,5	95,5	4,5	95,5	4,2	95,8
Ongevalsextsels en vergiftigingen	7,1	92,9	6,7	93,3	6,4	93,6
Fact. die de gezondh.toest. beïnv. en contacten met gezondh.diensten	4,1	95,9	4,0	96,0	3,6	96,4
Totaal	4,1	95,9	3,9	96,1	3,7	96,3

Hoofddiagnose

In de tabellen 20a en 20b worden de hoofddiagnose-groepen weergegeven naar koppelstatus en in de tijd. In tabel 20a worden de kolom-percentages gegeven (aandeel van de desbetreffende hoofddiagnose in het totaal van diagnoses bij de niet gekoppelde, resp. overige records) en in tabel 20b de rij-percentages (aandeel van niet gekoppelde records, resp. overige records in het totaal aantal records met de desbetreffende hoofddiagnose).

Uit tabel 20b blijkt dat het gemiddelde aandeel van de niet gekoppelde records bij de hoofddiagnosen psychische stoornissen (8,2%), ongevalsletsels (6,5%), infectieuze en parasitaire ziekten (6,0%), en in mindere mate bij aandoeningen van de perinatale periode (5,3%) en endocriene, voedings- en stofwisselingsziekten

(4,7%) relatief groot is ten opzichte van het gemiddelde percentage van 3,6% over alle hoofddiagnosen. In de tijd neemt het aandeel niet gekoppelde records over de hele linie af. De aandelen per hoofddiagnose nemen ook proportioneel af, behalve bij de psychische stoornissen: het aandeel van de niet gekoppelde records is hier in de tijd sterker afgenomen dan gemiddeld. In 1995 was het percentage niet gekoppelde records bij psychische stoornissen (11,7%) 2,6 maal zo groot dan gemiddeld, in 2001 was dit met 7,5% 2,3 maal zo groot.

Tabel 20a relateert de gevonden verschillen enigszins. De oververtegenwoordiging van bovengenoemde diagnoses komt ook hier naar voren, maar het aandeel van deze diagnoses in het totaal aantal opnamen is relatief beperkt. Zo maken de psychische stoornissen gemiddeld slechts 2,3%, resp. 1,0% uit van alle niet

1998		1999		2000		2001		Totaal	
niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld
1,5	1,0	1,5	0,9	1,8	1,0	1,6	0,9	1,6	1,0
6,7	8,6	6,1	8,8	6,5	9,1	6,6	9,3	6,5	8,7
1,2	1,0	1,2	1,0	1,2	1,0	1,1	1,0	1,1	1,0
1,9	1,5	2,0	1,6	2,0	1,6	2,1	1,6	2,1	1,6
2,3	1,0	2,2	1,0	2,0	1,0	2,1	0,9	2,3	1,0
6,7	9,5	6,4	9,3	6,7	9,5	7,1	9,6	6,7	9,4
11,2	11,9	10,3	11,7	10,1	11,6	10,8	11,3	10,9	11,9
7,6	7,9	7,4	7,2	7,1	6,9	6,9	6,7	7,4	7,5
7,7	7,3	7,4	7,3	7,7	7,4	7,8	7,3	7,8	7,4
1,5	1,3	1,3	1,3	1,5	1,4	1,7	1,5	1,5	1,4
7,0	10,1	6,8	10,0	6,7	9,7	7,1	9,9	7,3	10,3
5,2	5,9	5,1	5,7	4,9	5,6	5,2	5,4	5,3	6,0
6,4	6,5	7,0	6,5	7,2	6,6	7,5	6,5	6,6	6,4
4,3	2,7	4,4	2,8	4,6	2,9	2,8	2,9	4,0	2,7
1,0	1,0	1,1	1,0	1,0	1,0	1,0	1,0	1,1	1,0
6,8	5,9	7,7	6,5	6,8	6,0	7,2	6,0	6,6	5,7
9,7	5,3	9,9	5,4	10,0	5,3	10,2	5,2	10,0	5,5
11,2	11,5	12,0	12,0	12,2	12,6	11,2	12,9	11,2	11,6
84 698	2 225 653	81 340	2 249 657	79 651	2 253 183	79 494	2 339 265	587 106	15 528 506

1998		1999		2000		2001		Totaal	
niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld	niet-gekoppeld	gekoppeld
5,8	94,2	5,5	94,5	6,1	93,9	5,6	94,4	6,0	94,0
2,9	97,1	2,4	97,6	2,5	97,5	2,3	97,7	2,8	97,2
4,3	95,7	4,0	96,0	3,9	96,1	3,5	96,5	4,0	96,0
4,6	95,4	4,4	95,6	4,3	95,7	4,3	95,7	4,7	95,3
8,0	92,0	7,5	92,5	6,9	93,1	7,5	92,5	8,2	91,8
2,6	97,4	2,4	97,6	2,4	97,6	2,4	97,6	2,6	97,4
3,4	96,6	3,1	96,9	3,0	97,0	3,2	96,8	3,3	96,7
3,6	96,4	3,6	96,4	3,5	96,5	3,4	96,6	3,6	96,4
3,8	96,2	3,6	96,4	3,6	96,4	3,5	96,5	3,8	96,2
4,2	95,8	3,5	96,5	3,7	96,3	3,7	96,3	4,0	96,0
2,6	97,4	2,4	97,6	2,4	97,6	2,4	97,6	2,6	97,4
3,2	96,8	3,2	96,8	3,0	97,0	3,2	96,8	3,2	96,8
3,6	96,4	3,7	96,3	3,7	96,3	3,7	96,3	3,8	96,2
5,7	94,3	5,5	94,5	5,3	94,7	3,3	96,7	5,3	94,7
3,7	96,3	3,8	96,2	3,4	96,6	3,2	96,8	3,8	96,2
4,2	95,8	4,2	95,8	3,9	96,1	3,9	96,1	4,2	95,8
6,5	93,5	6,2	93,8	6,2	93,8	6,2	93,8	6,5	93,5
3,6	96,4	3,5	96,5	3,3	96,7	2,9	97,1	3,5	96,5
3,7	96,3	3,5	96,5	3,4	96,6	3,3	96,7	3,6	96,4

gekoppelde, resp. overige opnamen. Bij de infectieuze en parasitaire ziekten zijn de vergelijkbare percentages nog lager: 1,6% en 1,0%. Uit tabel 20a blijkt dat in grote lijnen de verdeling van diagnoses bij de niet gekoppelde en overige records geen grote verschillen vertoont, behalve bij de ongevalsletsels en vergiftigingen waarvan het aandeel in de niet-gekoppelde groep gemiddeld 10,0% is en bij de overige records 5,5%. Hier tikt de oververtegenwoordiging wat meer door op het totaal aantal niet gekoppelde records.

Conclusies niet gekoppelde records

Uit dit onderzoek naar de kenmerken van de niet gekoppelde records ten opzichte van de overige LMR-records komt naar voren

dat het hier gaat om een enigszins selectieve groep. Bij de niet-gekoppelden is de gemiddelde leeftijd iets lager (meer 0-jarigen, jongvolwassenen en zeer ouderen), betreft het vaker mannen, klinische opnamen, psychische stoornissen en ongevalsletsels en zijn de patiënten vaker afkomstig uit de grote steden in het westen des lands. De verschillen zijn weliswaar duidelijk waarneembaar in de groep niet gekoppelde records, maar aan de andere kant zijn de verschillen in de grote groep van wel gekoppelde records veel geringer.

In de tijd gezien neemt het totaal percentage niet-gekoppelden geleidelijk af van 4,1 in 1995 tot 3,3 in 2001. De gevonden verschillen in kenmerken vertonen geen grote variatie over de jaren, alleen neemt het leeftijdsverschil iets toe met de tijd en de oververtegenwoordiging van de psychische stoornissen wat af.

6. Conclusies kwaliteitsonderzoek

Bij het kwaliteitsonderzoek van de koppeling kan onderscheid gemaakt worden tussen miskoppelingen, d.w.z. onterecht gekoppelde records (foutpositieve koppelingen), en gemiste koppelingen, d.w.z. onterecht niet gekoppelde records (foutnegatieve koppelingen).

Administratieve fouten in de registraties zullen meestal leiden tot gemiste koppelingen, maar kunnen ook miskoppelingen tot gevolg hebben. * Soms is het ook mogelijk dat een administratieve fout toch tot een juiste koppeling leidt. Bijvoorbeeld, in het geval van de koppeling van de LMR met de GBA zal een administratieve fout in de 2 letters van de postcode in de meeste gevallen toch leiden tot een juiste koppeling (in stap II, op basis van de postcode-4). Minder vaak zal dit leiden tot een gemiste koppeling (meervoudig gekoppeld, maar LMR-record was wel uniek geweest op de correcte postcode-6; of postcode koppelt niet) of tot een miskoppeling (koppelt a.g.v. de onjuiste postcode met een verkeerde GBA-persoon met dezelfde geboortedatum en geslacht).

De omvang van de gemiste koppelingen kan niet met zekerheid vastgesteld worden. Wel kan uit een onderzoek naar de kwaliteit van het adresgegeven in de GBA (zie 2.2) afgeleid worden dat de administratieve foutenruis van het postcode-gegeven in de GBA tussen de 1 en 2% ligt. Verder is bij 0,4% van de geboortedata de dag en/of maand onbekend. Stel dat de gezamenlijke administratieve fout op postcode, geslacht en geboortedatum 2% van de GBA-records betreft en dat deze in de LMR even groot is (geen gegevens bekend), en aannemende dat deze fouten met name leiden tot gemiste koppelingen, dan zou bij 4% van de LMR-records sprake zijn van gemiste koppelingen. Deze schatting komt redelijk overeen met het bij de LMR/GBA koppeling gevonden percentage niet-gekoppelde records van 3,6%. En bij dit laatste percentage zullen ook opnamen zijn van personen die terecht niet koppelen omdat deze niet staan ingeschreven in de GBA. Zo beredeneerd is het gevonden percentage niet-gekoppelde records (en gemiste koppelingen) zeker niet hoger dan dat wat op grond van de administratieve foutenruis verwacht mag worden. Doordat het percentage laag is zal het effect op de statistische uitkomsten in het algemeen niet groot zijn, tenzij de niet gekoppelde records een *sterk* selectieve groep betreffen en de analyse met name ingaat op variabelen die de aard van de selectiviteit raken. Uit het kwaliteitsonderzoek is gebleken dat de LMR-records die niet konden worden gekoppeld met de GBA inderdaad een enigszins selectieve groep betreffen in vergelijking met de uniek en meervoudig gekoppelde records. De gevonden verschillen in kenmerken zijn echter ook weer niet zeer groot en niet sterk fluctuerend

in de tijd. Hoewel men zich bij de analyse van de gekoppelde gegevens bewust dient te zijn van deze verschillen, is het gezien de relatief beperkte mate van selectiviteit en omvang van de niet-gekoppelde groep de verwachting dat de invloed hiervan op de meeste analyses klein zal zijn, met name als deze gericht zijn op verbanden bij de totale groep van gekoppelde records. Echter, wanneer men zich met name richt op bepaalde moeilijk waarneembare groepen zullen deze in het algemeen moeilijk zichtbaar gemaakt kunnen worden uit administratieve gegevensbronnen, zeker bij koppeling van registraties. Het GSB zal hier op geen uitzondering zijn.

Aan de mogelijke omvang van de miskoppelingen, de onterecht gekoppelde records, is in dit hoofdstuk met name aandacht besteed. Uit vergelijking van de resultaten van koppeling op postcode-6 niveau met die op postcode-4 niveau van de LMR-records met een potentieel geldige postcode-6, is naar voren gekomen dat minder dan 3% van de uniek op postcode-4 niveau gekoppelde records miskoppelingen kunnen betreffen. Hoe groot het aandeel precies is is niet bekend, maar onderzoeken naar inconsistenties in het overlijdensgegeven en inconsistenties in de vorm van overlappende opnamen wijzen op lagere percentages miskoppelingen. Bij maximaal 1,5% van de in de LMR overleden patiënten is er een substantieel verschil in overlijdensdatum bij de aangekoppelde GBA-records; dit hoeven overigens niet altijd miskoppelingen te betreffen. En bij minder dan 0,1% van de uniek gekoppelde records is er sprake van overlappende opnamen die in de richting van miskoppeling kunnen wijzen. Samengevat zal het miskoppelingspercentage waarschijnlijk in de orde van grootte van 1% zijn en in ieder geval minder dan 3% bedragen. Op grond hiervan kan gesteld worden dat de miskoppelingen van beperkte omvang zijn.

Samengevat kan geconcludeerd worden dat, gezien het slechts in beperkte mate voorkomen van miskoppelingen en gemiste koppelingen, de in dit rapport beschreven koppeling van de LMR met de GBA van goede kwaliteit is en een voldoende basis vormt voor statistische analyses.

* Er van uitgaande dat er 85 (leeftijd) x 365 (dagen) x 10^4 (postcode 4 pos.) x 2 (geslacht) = 620×10^6 verschillende waarden kunnen zijn van de geboortedatum/postcode/geslacht sleutel, en uitgaande van een populatie in Nederland van 16×10^6 , dan is de kans op een miskoppeling bij een administratieve fout ca. 1 : 40.

7. Beschouwing

Bij de in dit rapport beschreven koppeling van LMR- met GBA-gegevens is uitgegaan van de primaire koppelvariabelen opnamedatum, geboortedatum, geslacht en postcode (eerst op 6 posities in stap I, dan op 4 posities in stap II), en daarna is ook de overlijdensdatum van in het ziekenhuis overleden LMR-patiënten als koppelvariabele meegenomen (stap III).

Er is voor gekozen om bij stap II (postcode-4 koppeling) te koppelen met de gehele GBA-populatie, en hier dus niet de GBA-persoon buiten beschouwing te laten die al gekoppeld zijn in stap I (postcode-6 koppeling). In theorie zou door het 'wegstrepen' van de al eerder gekoppelde GBA-persoon bij de meervoudig-gekoppelden in stap II, een verhoging van het koppelrendement bereikt kunnen worden. Dit zou echter alleen toegepast mogen worden in geval van overlappende opnameperioden bij het in stap I gekoppelde LMR-record en het desbetreffende LMR-record in stap II, omdat er heropnamen kunnen plaatsvinden bij dezelfde persoon, die dan door het 'wegstrepen' van deze persoon onterecht buiten beschouwing zouden worden gelaten. Uit het kwaliteitsonderzoek van de nu uitgevoerde koppeling is echter gebleken dat zelfs bij overlappende opnamen deze opnamen in een aanzienlijk aantal gevallen toch toegewezen kunnen worden aan dezelfde persoon. Mede daarom is afgezien van deze extra slag in de koppeling.

De nu uitgevoerde koppeling kan in principe nog op verschillende manieren worden uitgebreid:

- Er kan nog een extra koppelingsslag gedaan worden waarbij alleen gekoppeld wordt op geboortedatum, geslacht en overlijdensdatum. De extra koppelingen die dan worden bereikt betreffen dan in het ziekenhuis overleden patiënten waarvan de postcode niet koppelt met de GBA. Gebleken is dat met een dergelijke koppeling ca. 13 000 LMR-records van de eerder niet gekoppelde records alsnog uniek gekoppeld kunnen worden (1995–2001); op het totaal aantal records is dit +0,08%. Voor analyses waarbij de sterfte niet als variabele wordt meegenomen zal deze aanvulling niet van belang zijn. Bij analyses waarbij een zo volledig mogelijk beeld van de ziekenhuissterfte gewenst is, kan eventueel overwogen worden om deze extra koppelingsslag wel mee te nemen.

De selectiviteit van de niet-gekoppelde records verandert, uitgaande van de variabelen zoals beschreven in paragraaf 5.7, niet veel door deze extra koppelingsslag. Wel zijn er vanzelfsprekend aanzienlijk minder in het ziekenhuis overleden

patiënten in deze groep (0,8%) in vergelijking met de niet-gekoppelden vóór de extra koppelingsslag (3,0%). Bij de grote groep overige LMR-records verandert dit percentage nauwelijks door de extra koppelingsslag (van 2,1% naar 2,0%)

- Verder kan een deel van de nu meervoudig en niet gekoppelde records nog worden 'bijgekoppeld' bij de uniek-gekoppelden op basis van overeenkomstige waarden op secundaire koppelingsvariabelen. Het gaat hier bijvoorbeeld om de variabelen patiëntnummer en huisartscode in de LMR. Als in de loop der jaren één of meer opnamerecords van een persoon uniek gekoppeld zijn op de primaire koppelingsvariabelen, dan kunnen nog niet uniek gekoppelde records van dezelfde persoon (bijvoorbeeld als gevolg van wijziging van de uniciteit door verhuizing) worden bijgekoppeld doordat naast de geboortedatum en geslacht, patiëntnummer en/of huisartscode overeenkomen. Deze methode kan echter alleen met enige beperkingen worden toegepast omdat de patiëntnummers en huisartscodes meestal ziekenhuis-specifiek zijn en patiëntnummers niet altijd patiënt-specifiek worden toegekend. Daarnaast kunnen wellicht ook LMR-records worden bijgekoppeld door koppeling met andere registraties waarin gegevens over ziekenhuisopnamen zijn vastgelegd.

Deze aanvullende koppelingsmogelijkheden zijn nog onderwerp van studie. In het algemeen zullen 'bijkoppelingen' pas in de productie worden toegepast als hiermee het merendeel van de nu niet uniek gekoppelde records kan worden gekoppeld. Anders wordt de ophoging van de gegevens zeer complex, omdat de koppelingen van bijgekoppelde records en de al eerder gekoppelde records op zeer verschillende wijzen tot stand zijn gekomen.

Voor de records die uniek gekoppeld zijn volgens de procedure van de in dit rapport beschreven koppeling wordt nu door het CBS een ophoogmethodiek ontwikkeld. Door te corrigeren voor de verschillende koppelkansen (als gevolg van de verschillen in uniciteit op de primaire koppelvariabelen) kan op opnameniveau gecorrigeerd worden voor de meervoudig-gekoppelden. Ook is het mogelijk om op opnameniveau op te hogen voor de niet in de LMR geregistreerde opnamen, voor zover hier randtotalen van beschikbaar zijn (zie hoofdstuk 3), en eventueel voor niet gekoppelde LMR-records. Voor ophoging op persoonsniveau is dit complexer, ook door veranderingen in de uniciteit van een persoon in de tijd.

Bijlagen

Bijlage 1. Inhoud LMR-bestanden

Bestand opos

Omschrijving	Kolomnaam	Null?	Type kolom	Fac ¹⁾	Lengte
Bestandsidentificator	DEGNAAM		VARCHAR2		8
Registratiejaar	DEREGJR	NOT NULL	VARCHAR2		2
Instellingsnummer	DEINSTN	NOT NULL	VARCHAR2		5
Locatienummer	DELCFD		VARCHAR2	X	1
Opnamenummer	DEOPNNR	NOT NULL	VARCHAR2		10
Heropname	DEHEROP		NUMBER		3
Type zorgverlening	DEZORTP		VARCHAR2		1
Opnamedatum	DEDATOP(dag)		DATE		8
Opname-uur	DEDATOP(uur)		VARCHAR2		2
Opnamereden	DEOPRED		NUMBER		3
Opname-urgentie	DEOPURG		NUMBER		3
Herkomst	DEHERKO		NUMBER		3
Herkomst instelling	DEHERKI		VARCHAR2		5
Periodenr hoofddiagnose	DEVAWHD		NUMBER		4
Periodenr hoofdverrichting	DEVAWHV		NUMBER		4
Ontslaguur	DEDATON		VARCHAR2		2
Bestemming	DEBSTMC		NUMBER		3
Bestemming-instelling	DEBSTMI		VARCHAR2		5
Obductie	DEOBDUC		VARCHAR2		1
Patiëntnummer	DEPATNR		VARCHAR2	X	10
Geslacht	DEGESLA		NUMBER		3
Geboortedatum	DEGEBDT		DATE		8
Woongemeente (=postcode 4 pos.)	DEPOSCE		VARCHAR2		4
Wijk (meestal 2 letters postcode)	DEWIJCE		VARCHAR2	X	2
Huisarts	DEHRTCE		VARCHAR2		5
Verzekeringwijze	DEVZWYZ		VARCHAR2		1
Betalende instantie	DEBTICE		VARCHAR2	X	6
Voorzorg-zorgverlener (discipline)	DEVDISC		VARCHAR2	X	2
Vorzorg-zorgverlener (specialisme)	DEVSPME		VARCHAR2	X	2
Voorzorginstantie	DEVINTA		VARCHAR2		2
Voorzorginstelling	DEVINST		VARCHAR2	X	5
Voorzorgtype	DEVTYPE		VARCHAR2		1
Nazorg-zorgverlener (discipline)	DENDISC		VARCHAR2	X	2
Nazorg-zorgverlener (specialisme)	DENSPME		VARCHAR2	X	2
Nazorginstantie	DENINTA		VARCHAR2		2
Nazorginstelling	DENINST		VARCHAR2	X	5
Nazorgtype	DENTYPE		VARCHAR2		1
Verpleegduur	DEWVPLD		NUMBER		6
Opnamedag	DEDATOP		VARCHAR2		2
Leeftijd	DEDATOP en DEGEBDT		VARCHAR2		3
Dag hoofdoperatie	DEDAGHV		VARCHAR2		2
Gemeente	DRGEMCE		NUMBER		4
Land	DRLANCE		VARCHAR2		2

Bestand diagnoses

Omschrijving	Kolomnaam	Null?	Type kolom	Fac ¹⁾	Lengte
Bestandsidentificator	DEGNAAM		VARCHAR2		8
Registratiejaar	DEREGJR	NOT NULL	VARCHAR2		2
Instellingsnummer	DEINSTN	NOT NULL	VARCHAR2		5
Opnamenummer	DEOPNNR	NOT NULL	VARCHAR2		10
Verantwoordelijkheidsperiodenummer	DEVAVPN	NOT NULL	NUMBER		4
Datum overdracht	DEDATOV		DATE		8
Indicatie vervolg	DEIVERV		NUMBER		3
Werkelijke preoperatieve verpleegduur	DEWPVPD		NUMBER		6
Diagnosenummer	DEDIAGN	NOT NULL	NUMBER		4
Primaire diagnose	DEPDIAG		VARCHAR2		1
Emc code	DEEMCCE		VARCHAR2		1
Diagnose code 5	DEDIAG5		VARCHAR2		5
Specialisme	DESPMCE		VARCHAR2		2

¹⁾ Facultatief gegeven, d.w.z. niet alle instellingen registreren deze variabele.

Bestand verrichtingen

Omschrijving	Kolomnaam	Null?	Type kolom	Fac ¹⁾	Lengte
Bestandsidentificator	DEGNAAM		VARCHAR2		8
Registratiejaar	DEREGJR		VARCHAR2		2
Instellingsnummer	DEINSTN	NOT NULL	VARCHAR2		5
Opnamenummer	DEOPNNR	NOT NULL	VARCHAR2		10
Verantwoordelijkheidsperiodenummer	DEVAWPN	NOT NULL	NUMBER		4
Primaire verrichting	DEPVERR		VARCHAR2		1
Verrichting nummer	DEVERRN	NOT NULL	NUMBER		2
Verrichtingcode	DEVERR6		VARCHAR2		6
Diagnosenummer	DEDIAGN		NUMBER		2
Datum verrichting	DEDAVER		DATE		8
Instelling verrichting	DEINVER		VARCHAR2		5
Code specialisme als elders	DESPMCE		VARCHAR2		2

Bestand verantwoordelijk specialisme

Omschrijving	Kolomnaam	Null?	Type kolom	Fac ¹⁾	Lengte
Bestandsidentificator	DEGNAAM		VARCHAR2		8
Registratiejaar	DEREGJR		VARCHAR2		2
Instellingsnummer	DEINSTN	NOT NULL	VARCHAR2		5
Opnamenummer	DEOPNNR	NOT NULL	VARCHAR2		10
Verantwoordelijkheidsperiodenummer	DEVAWPN	NOT NULL	NUMBER		4
Specialismenummer verantwoordelijk specialist	DERSPMCE	NOT NULL	VARCHAR2		2

Bestand specialisme operaties (indien operatie in andere instelling)

Omschrijving	Kolomnaam	Null?	Type kolom	Fac ¹⁾	Lengte
Bestandsidentificator	DEGNAAM		VARCHAR2		8
Registratiejaar	DEREGJR		VARCHAR2		2
ararsid9718680 Instellingsnummer	DEINSTN	NOT NULL	VARCHAR2		5
Opnamenummer	DEOPNNR	NOT NULL	VARCHAR2		10
Verantwoordelijkheidsperiodenummer	DEVAWPN	NOT NULL	NUMBER		4
Verrichting nummer	DEPVERRN	NOT NULL	NUMBER		2
Specialisme	DESPMCE		VARCHAR2		2

¹⁾ Facultatief gegeven, d.w.z. niet alle instellingen registreren deze variabele.

Bijlage 2. Bij koppeling toegevoegde gegevens

Bij de koppeling van de LMR-records aan het CKP worden de volgende variabelen toegevoegd:

– **RINPersoon**

Willekeurig nummer in het CKP, toegekend aan een persoon; elk nummer is uniek.

– **RINPersoonVolgnr**

Bij elke wijziging in de persoonsvariabelen (zoals burgerlijke staat en nationaliteit) is een nieuw record aangemaakt in het CKP. Elk record is voorzien van een volgnummer. Dit geldt niet voor adreswijzigingen, hier wordt het volgnummer niet opgehoogd. RINPersoon gecombineerd met RINPersoonVolgnr identificeert een persoonsrecord in het CKP.

– **GBAPostcode (6 posities)**

Wordt voor analysedoeleinden aan gekoppelde records toegevoegd.

– **CBKSoortNr**

Geeft aan of er gekoppeld is of niet. G = geen koppeling, R = wel koppeling.

– **CBKBeginGeldigheid**

Begindatum geldigheid van het record.

– **CBKEindGeldigheid**

Einddatum geldigheid van het record.

– **CBKAfstand**

Afstand in dagen tussen geldigheidsperiode aangekoppelde CKP-record en de referentiedatum van het bronrecord (hier: LMR-record).

– **Na**

Aantal records in bronbestand met gelijke primaire koppelsleutelwaarden.

– **NofA**

Volgnummer binnen de Na records.

– **Nb**

Aantal aangekoppelde CKP-records met gelijke primaire koppelsleutelwaarden.

– **NofB**

Volgnummer binnen de Nb records.

– **CBKMutatiecode**

Code die aangeeft welke soort of soorten mutaties in de GBA heeft/hebben plaatsgevonden. Geeft aan waarom het desbetreffende record is beëindigd of gestart. De code ontstaat door de waarden van de verschillende mutaties op te tellen (bereik: 0 .. 8191):

0	geen wijziging
1	wijziging in geslacht
2	wijziging in nationaliteit, tweede nationaliteit of type nationaliteit
4	wijziging in verblijfstitel
8	wijziging in burgerlijke staat
16	wijziging in familierechterlijke betrekkingen moeder
32	wijziging in familierechterlijke betrekkingen vader
64	opvoer = geboorte
128	opvoer = immigratie
256	overige opvoer
512	afvoer = sterfte
1024	afvoer = emigratie
2048	overige afvoer
4096	adreswijziging

– **CBKModel**

Samengestelde code waaruit afgeleid kan worden welke koppelingsstappen zijn doorlopen:

1	afkomstig uit koppeling op postcode-6
2	afkomstig uit koppeling op postcode-4
3	uniek op zowel postcode-6 als postcode-4
14	meervoudig bij koppeling op postcode-6, vervolgens door overlijdensfilter
24	meervoudig bij koppeling op postcode-4, vervolgens door overlijdensfilter

Bijlage 3. Resultaten overlappende opnamen bij dezelfde instelling

In vervolg op wat bij paragraaf 5.6 is beschreven over het vóórkomen van overlappende opnamen met andere instellingen, is een zelfde soort onderzoek gedaan naar overlappende opnamen die betrekking hebben op dezelfde instelling. Het gaat om totaal 5 042 records, dat wil zeggen bij 2 521 record-paren. De resultaten van de bestemmingsvariabele staan in tabel 1.

Hieruit blijkt een ander beeld dan bij de overlappende opnamen in een andere instelling. Bij overlappende opnamen in dezelfde instelling maakt het geen verschil of er onderscheid gemaakt wordt tussen tweede opnamen die wel of niet doorlopen na de eerste opnamen. De overgrote meerderheid van de eerste opnamen (92%), heeft de eigen woonomgeving als bestemming. Dat 'andere instelling' hier erg laag scoort is, uitgaande van de veronderstelling dat de overlappende opnamen inderdaad toebehoren tot dezelfde persoon, te verwachten. Omdat de waarde 'zelfde instelling' in de LMR-registratie niet bij de variabele bestemming kan worden ingevuld, lijkt het aannemelijk dat hiervoor dan dezelfde waarde wordt ingevuld als bij de tweede opname, d.w.z. meestal de 'eigen woonomgeving'. Wanneer toch 'andere instelling' wordt ingevuld, blijkt dit, volgens het ingevulde instellingsnummer, in een kwart van de gevallen toch dezelfde instelling te zijn.

De uitkomsten bij de herkomst variabele staan vermeld in tabel 2.

Tabel 2
Herkomst van tweede opname bij overlap met eerste opname in zelfde instelling, bij 2 521 paren van records (1995–2001)

	Abs.	%	
Herkomst			
Eigen woonomgeving	1 149	45,6	
Bejaardenoord	6	0,2	
Andere instelling	72	2,9	
w.v. % 'juist' nr. 1)			44%
w.v. % 'onjuist' nr. 1)			56%
In ziekenhuis geboren	1 294	51,3	
Totaal	2 521	100,0	

1) Percentage van de records met ingevuld instellingsnummer waarbij het instellingsnummer van de herkomst overeenkomt ('juist' nr.), resp. niet overeenkomt ('onjuist' nr.) met het instellingsnummer van de 1e (en in dit geval ook 2e) opname.

Ook hier blijkt dat 'andere instelling' laag scoort. Verder valt op dat naast de 'eigen woonomgeving' ook 'in ziekenhuis geboren' relatief hoog scoort als herkomst. Dit wijst erop dat de overlappende opnamen in dezelfde instelling relatief vaak pasgeborenen betreft.

Tabel 1
Bestemming na eerste opname bij overlap met tweede opname in dezelfde instelling, bij 2 521 paren van records (1995–2001)

Bestemming	Totaal		2e opname valt binnen 1e opn. periode		2e opname eindigt na 1e opname		
	abs.	%	abs.	%	abs.	%	
Eigen woonomgeving	2 315	91,8	1 190	91,0	1 125	92,7	
Bejaardenoord	10	0,4	8	0,6	2	0,2	
Andere instelling	117	4,6	72	5,5	45	3,7	
w.v. % 'juist' nr. 1)							26%
w.v. % 'onjuist' nr. 1)							74%
In instelling overleden	76	3,0	35	2,7	41	3,4	
Tegen advies weggegaan	3	0,1	2	0,2	1	0,1	
Totaal	2 521	100,0	1 307	100,0	1 214	100,0	

1) Percentage van de records met ingevuld instellingsnummer waarbij het instellingsnummer van de bestemming overeenkomt ('juist' nr.), resp. niet overeenkomt ('onjuist' nr.) met het instellingsnummer van de 2e (en in dit geval ook 1e) opname.

Referenties

- 1) Arts C.H., Hoogteijling E.M.J. Het Sociaal Statistisch Bestand 1998 en 1999. Sociaal-Economische Maandstatistiek 2002; 19(12); 13–21.
- 2) Stichting Prismant en Centraal Bureau voor de Statistiek. Overeenkomst verkrijging en verstrekking van gegevens. CBS archiefnr. 325-02-SRP. Voorburg, 20 februari 2002.
- 3) SIG Zorginformatie. Gebruikershandboek Landelijke Medische Registratie. Versie 03.93. Utrecht: Prismant, 1993.
- 4) Paas R.A., Veenhuizen, C.W. Onderzoek naar de betrouwbaarheid van de LMR. Rapportage voor de ziekenhuizen. Utrecht: Prismant, februari 2002.
- 5) Prins C.J.M. Dutch population statistics based on population register data. Maandstatistiek van de Bevolking 2000; 48(2): 9–15.
- 6) Basisadministratie Persoonsgegevens en Reisdocumenten. De kwaliteit van het adresgegeven in de GBA. 's-Gravenhage: Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, Agentschap BPR, september 2002.
- 7) CBS, Prismant & Afdeling Klinische Epidemiologie & Biostatistiek van het AMC. Het herkennen van ziekenhuisopnamen in cohort-onderzoek via landelijke registraties. Voorburg: CBS, augustus 2001.