

# **HET HERKENNEN VAN ZIEKENHUISOPNAMEN EN OVERLIJDEN IN COHORT-ONDERZOEK VIA LANDELIJKE REGISTRATIES**

Een pilotstudie naar de technische uitvoerbaarheid en de kwaliteit van het anoniem koppelen van records uit de Landelijke Medische Registratie, de Gemeentelijke Basisadministratie en de Statistiek van Doodsoorzaken

*Een samenwerkingsproject van:*

*Centraal Bureau voor de Statistiek*

*Prismant*

*Afdeling Klinische Epidemiologie & Biostatistiek van het AMC*

*mogelijk gemaakt door een subsidie van de Nederlandse Hartstichting*

*Augustus 2001*

**DE VOLGENDE PERSONEN EN INSTELLINGEN HEBBEN AAN DIT ONDERZOEK  
BIJGEDRAGEN:**

**CENTRAAL BUREAU VOOR DE STATISTIEK**

Agnes de Bruin  
Jan Kardaun  
Peter Barendswaard  
Kees Prins

Contactadres: Centraal Bureau voor de Statistiek \*  
Postbus 4000  
2270 JM Voorburg

**PRISMANT (VOORHEEN SIG ZORGINFORMATIE)**

Evelien Gevers  
Johan van der Wal

Contactadres: Prismant  
Postbus 85200  
3508 AE Utrecht

**AFDELING KLINISCHE EPIDEMIOLOGIE & BIostatistiek  
ACADEMISCH MEDISCH CENTRUM**

Gouke Bonsel (thans Sociale Geneeskunde, Academisch Medisch Centrum)  
Hans Reitsma (tevens Nederlandse Hartstichting)

Contactadres: Afdeling Klinische Epidemiologie & Biostatistiek, AMC  
Postbus 22660  
1100 DD Amsterdam

**NEDERLANDSE HARTSTICHTING**

Ineke van Dis (waarnemer)

Contactadres: Nederlandse Hartstichting  
Postbus 300  
2501 CH Den Haag

---

\* Exemplaren van dit rapport kunnen worden aangevraagd bij het Centraal Bureau voor de Statistiek, Voorburg, tel. (070) 337 5299

# INHOUDSOPGAVE

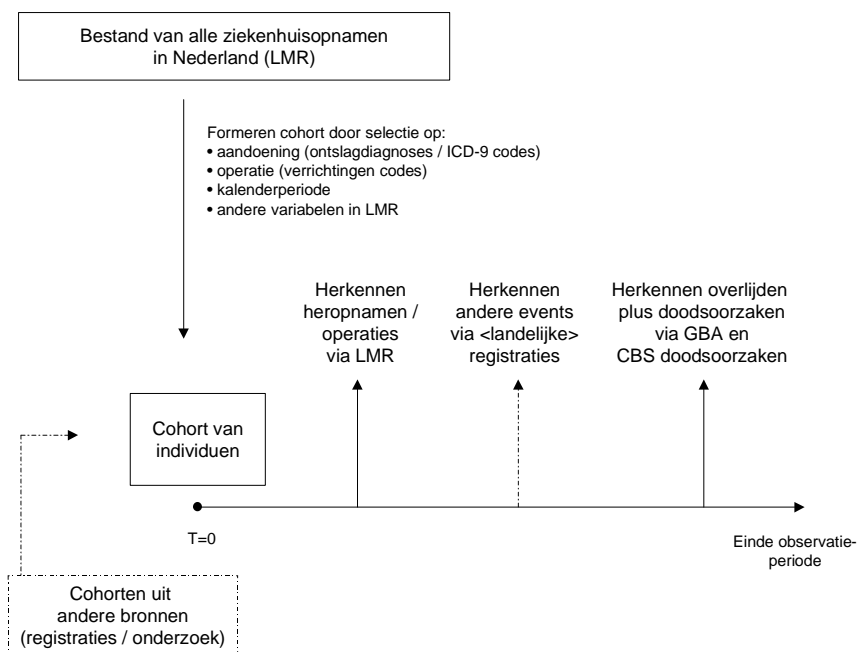
1.	INLEIDING	3
2.	HERKENNEN OVERLIJDEN EN HEROPNAMEN VIA LANDELIJKE REGISTRATIES	7
	2.1 Herkennen van overlijden	7
	2.2 Herkennen van heropnamen	7
3.	DOEL PILOT EN UITGEVOERDE ACTIVITEITEN	9
	3.1 Kwaliteit initiële koppeling tussen LMR en GBA	9
	3.1.1 uniciteit primaire koppelsleutel in GBA	
	3.1.2 uniciteit primaire koppelsleutel in LMR-cohort	
	3.1.3 GBA koppelstatus van patiënten uit LMR-cohort	
	3.1.4 patroon van overlijden in LMR-cohort	
	3.1.5 verkeerde koppelingen tussen LMR en GBA records	
	3.2 Meerwaarde GBA voor herkennen heropnamen	11
	3.2.1 verhuispatroon alle Nederlanders	
	3.2.2 verhuispatroon LMR-cohort	
	3.2.3 verlies unieke GBA koppelstatus in de tijd	
	3.2.4 heropnamefrequentie in relatie tot verhuizen	
4.	RESULTATEN MET KORTE BESPREKING	13
	4.1 Resultaten kwaliteit initiële koppeling tussen LMR en GBA	13
	4.1.1 uniciteit primaire koppelsleutel in GBA	
	4.1.2 uniciteit primaire koppelsleutel in LMR-cohort	
	4.1.3 GBA koppelstatus van patiënten uit LMR-cohort	
	4.1.4 patroon van overlijden in LMR-cohort	
	4.1.5 verkeerde koppelingen tussen LMR en GBA records	
	4.2 Resultaten meerwaarde GBA voor herkennen heropnamen	20
	4.2.1 verhuispatroon alle Nederlanders	
	4.2.2 verhuispatroon LMR-cohort	
	4.2.3 verlies unieke GBA koppelstatus in de tijd	
	4.2.4 heropnamefrequentie in relatie tot verhuizen	
5.	DISCUSSIE EN CONCLUSIES	27
	5.1 Keuzes bij opzet cohort-onderzoek	27
	5.2 Herkennen van overlijden	29
	5.3 Herkennen van heropnamen	32
6.	SAMENVATTING	35
BIJLAGE		
	Formules voor het aantal administratieve meerlingen	36



# 1. INLEIDING

Gegevens uit landelijke registraties worden frequent gebruikt om de gevolgen van ziekten op nationaal niveau te beschrijven. Ook het project *'Hart- en vaatziekten in Nederland, een epidemiologische verkenning'* van de Nederlandse Hartstichting maakt gebruik van landelijke cijfers over ziekenhuisopnamen en sterfte om het belang van hart- en vaatziekten in Nederland zichtbaar te maken. De huidige analyses kenmerken zich door jaar-op-jaar analyses per gegevensbron.

Analyses waarbij individuele patiënten in de tijd worden gevolgd of analyses waarbij bronnen in samenhang worden geanalyseerd, zijn schaars. Dit is jammer omdat het belang van longitudinale gegevens steeds groter wordt, onder meer voor het beschrijven van de lotgevallen van patiënten die door moderne behandelingsmethoden in leven blijven. Binnen de hart- en vaatziekten heeft dit geleid tot een duidelijke verschuiving van acute naar meer chronische vormen van hart- en vaatziekten. Dit betekent ook dat de aandacht binnen epidemiologisch onderzoek zich verplaatst van sterfte naar ziekte, en van korte-termijn gevolgen naar gezondheidseffecten op de lange duur. Langlopende nationale registraties kunnen in deze ontwikkeling een belangrijke rol spelen, indien patiënten binnen en tussen registraties gevolgd kunnen worden. Figuur 1 geeft een schematisch overzicht van het volgen van patiënten via landelijke registraties.

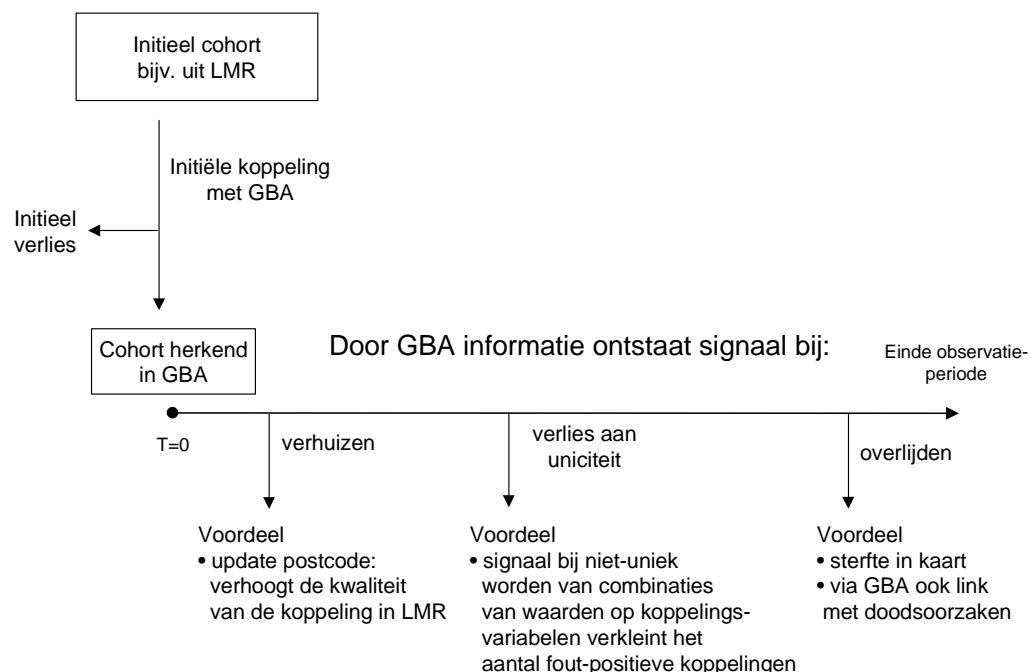


**Figuur 1.** Algemene opzet van cohort-onderzoek waarbij 'events' worden herkend via bestaande (landelijke) registraties. De gestippelde onderdelen zijn niet in de pilot onderzocht. LMR=Landelijke Medische Registratie; GBA=Gemeentelijke Basis Administratie.

Tot op heden zijn de mogelijkheden in Nederland voor dergelijk onderzoek gering. Zo is de eenheid van registreren in de Landelijke Medische Registratie (LMR) van Prismant (nieuwe naam na samengaan SIG Zorginformatie met Nationaal Ziekenhuis Instituut) een opname in plaats van een persoon. Dit betekent dat bij de (standaard) analyse van LMR-gegevens heropnamen verborgen blijven. Het ontbreken van een nationaal identificatienummer in de gezondheidszorg maakt dat gegevens uit individuele registraties moeilijk in samenhang kunnen worden geanalyseerd.

### *Nieuwe ontwikkelingen door komst Gemeentelijke Basisadministratie*

De komst van de Gemeentelijk Basisadministratie (GBA) in 1995 vergroot in technische zin de mogelijkheden om in Nederland personen binnen en tussen registraties te volgen. De GBA bevat informatie over alle in Nederland ingeschreven personen. De GBA legt vele administratieve gegevens vast, onder meer het huidige woonadres. Het Centraal Bureau voor de Statistiek (CBS) heeft ten behoeve van haar statistische taak de beschikking over structuurgegevens (stand van de bevolking per 1 januari) en stroomgegevens (mutaties door onder meer migratie, geboorte en sterfte). Met behulp van deze structuur- en stroomgegevens kan het CBS de reguliere bevolkingsstatistieken produceren en tevens voor steekproeven uit de bevolking de demografische gegevens actualiseren in de tijd. De GBA kan drie belangrijke functies vervullen bij het volgen van personen in de tijd (zie figuur 2).



**Figuur 2.** Centrale rol van de Gemeentelijke Basisadministratie (GBA) bij het combineren van gegevens uit verschillende registraties, in het bijzonder uit de Landelijke Medische Registratie en de Statistiek van Doodsoorzaken.

Ten eerste, de GBA registreert het overlijden van personen. Dit biedt de mogelijkheid om binnen een groep personen, eenmaal herkend in de GBA, het patroon van overlijden te beschrijven (passieve follow-up). Door een directe koppeling van GBA met CBS-doodsoorzakenstatistiek is tevens een analyse mogelijk naar primaire en secundaire doodsoorzaken. Ten tweede, de GBA registreert verhuizingen van individuen, inclusief het nieuwe adres. Dit biedt de mogelijkheid om de kwaliteit van koppelingen te verhogen binnen en tussen registraties waarbij gebruik gemaakt moet worden van variabelen die samenhangen met het huidige woonadres, zoals postcode en straatnaam. Ten derde, de GBA signaleert wanneer een persoon met een voorheen unieke combinatie van waarden op bepaalde koppelingsvariabelen niet meer uniek is en andersom. Ook deze informatie kan de kwaliteit koppelingen verhogen binnen en tussen registraties die gebruik maken van deze variabelen, in het bijzonder door het vermijden van het koppelen van records die behoren tot de administratieve wederhelft in plaats van de echte cohortdeelnemer.

Daarnaast bevat de GBA zelf informatie die van belang kan zijn. Zo is in de GBA onder meer informatie opgeslagen over de nationaliteit en het geboorteland van zowel de persoon zelf, als van zijn of haar ouders.

Door deze ontwikkelingen heeft de Nederlandse Hartstichting (NHS) het CBS en Prismant benaderd om de mogelijkheden na te gaan voor een (anoniem) cohortonderzoek van in het ziekenhuis opgenomen patiënten met een hart- of vaatziekte (zie ook brief NHS van 13 mei 1997 kenmerk IvD/97.0241). De huidige plannen zijn gebaseerd op het volgen van een cohort patiënten met een specifieke ontslag- en/of operatiecode (bijvoorbeeld acuut hartinfarct). De patiënten uit een dergelijk cohort worden gedurende een bepaalde periode gevolgd teneinde het patroon van heropnamen en overlijden vast te stellen (zie ook figuur 1). Het gaat hierbij zowel om de frequentie, de timing, alsmede oorzaken en kenmerken van opnamen en overlijden. Dit onderzoek lijkt uitvoerbaar door het combineren (koppelen) van Prismantgegevens (LMR) met gegevens uit de GBA en de doodsoorzakenregistratie van het CBS.

Het in samenhang analyseren van ziekte- en sterftegegevens biedt meerdere mogelijkheden, waaronder:

- ◆ Het in kaart brengen van de morbiditeit en mortaliteit van specifieke patiëntengroepen op nationaal niveau
- ◆ Het beschrijven van relaties tussen ziektebeelden; bijvoorbeeld de kans op (klinisch manifest) hartfalen na een myocardinfarct
- ◆ Het verbeteren van schattingen van de incidentie van bepaalde aandoeningen en het onderscheid kunnen maken tussen eerste en recidief gebeurtenissen
- ◆ De resultaten gebruiken als input-gegevens bij kosten van ziekten studies en scenario-analyses

Een pilotstudie werd noodzakelijk geacht om zowel de technische uitvoerbaarheid als de kwaliteit van de verschillende koppelingen te onderzoeken. De resultaten van

de pilot geven inzicht in de haalbaarheid van grootschalig cohortonderzoek in Nederland via registraties. Daarnaast dienen de resultaten als onderbouwing voor de diverse keuzen (zowel technisch als inhoudelijk), die moeten plaatsvinden tijdens een hoofdonderzoek. In de pilotstudie zijn proefonderzoeken uitgevoerd waarbij LMR- en GBA-gegevens zijn gekoppeld en vervolgens geanalyseerd. In dit verslag bespreken wij de mogelijkheden en problemen die wij tijdens verschillende koppelingen hebben ondervonden. De samenstelling van de projectgroep, die alle activiteiten heeft uitgevoerd en besproken, is terug te vinden in bijlage A.

Conform de huidige wet- en regelgeving zijn *alle* hier beschreven koppelingen en analyses op en door het CBS uitgevoerd. Tussen Prismant, CBS en NHS is hiertoe een tripartite overeenkomst gesloten, waarin gebruiksdoel en voorwaarden van de gegevensbestanden in detail zijn vastgelegd. Ook toekomstig onderzoek zal onder vergelijkbare voorwaarden kunnen plaatsvinden, waarbij de mogelijkheid wordt ingebouwd dat een 'externe' onderzoeker op het CBS analyses kan uitvoeren. Fundamenteel punt in de privacybescherming is dat alle resultaten, tabellen en figuren die het CBS verlaten niet tot een persoon/zorgverlener/instelling zijn te herleiden. Om dit punt te onderstrepen is in de subtitel van deze rapportage het woord 'anoniem' opgenomen.

Het rapport kent de volgende opbouw. Allereerst volgt een globale beschrijving van de typen koppelingen die nodig zijn voor het herkennen van heropnamen en overlijden, en de potentiële problemen die daarbij kunnen optreden (hoofdstuk 2). Vervolgens beschrijven wij in hoofdstuk 3 de specifieke problemen die in de pilot centraal staan en de activiteiten die zijn uitgevoerd om hier beter zicht op te krijgen. In hoofdstuk 4 worden de resultaten beschreven en kort toegelicht. De rapportage eindigt met een algehele bespreking van de bevindingen en formuleert aanbevelingen over toekomstig NHS-cohort onderzoek (hoofdstuk 5).



## **2. HERKENNEN OVERLIJDEN EN HEROPNAMEN VIA LANDELIJKE REGISTRATIES**

In de plannen van de Nederlandse Hartstichting staat het analyseren van het patroon van heropnamen en overlijden na een eerdere opname wegens een specifieke ziekte dan wel ingreep centraal. We schetsen allereerst in algemene zin de mogelijkheden en problemen bij het herkennen van overlijden (paragraaf 2.1) en heropnamen (paragraaf 2.2) via registraties.

### **2.1 HERKENNEN VAN OVERLIJDEN**

Patiënten uit een cohort die overlijden kunnen langs twee wegen worden herkend:

- a. via de LMR
- b. via de GBA

Beide methoden worden kort besproken.

#### *ad a. Herkennen overlijden via LMR gegevens*

Deze weg kent de volgende problemen:

- ◆ signaleert alleen de sterfte in het ziekenhuis
- ◆ mist sterfte in ziekenhuis als de bijbehorende (her)opname niet wordt herkend
- ◆ geeft geen informatie over doodsoorzaken, wel over opname- en ontslagdiagnosen. Deze kunnen echter afwijken van de doodsoorzaak

#### *ad b. Herkennen overlijden via GBA gegevens*

De GBA legt vast wanneer iemand overlijdt zowel binnen als buiten het ziekenhuis. Daarnaast is er een directe link (koppeling via gemeenschappelijk en uniek identificatienummer) met de statistiek van doodsoorzaken, waardoor informatie beschikbaar komt over primaire en secundaire doodsoorzaken. Mogelijke problemen bij deze route zijn:

- ◆ een persoon uit het LMR-cohort kan niet worden teruggevonden in de GBA
- ◆ een persoon uit het LMR-cohort koppelt met meerdere personen uit de GBA
- ◆ een persoon uit het LMR-cohort koppelt met één persoon uit de GBA, maar met de verkeerde
- ◆ een persoon overlijdt volgens de GBA, maar er is geen bijbehorend record in de doodsoorzakenstatistiek. Bijvoorbeeld voor Nederlanders die in het buitenland overlijden, wordt meestal geen doodsoorzakenformulier ontvangen
- ◆ een persoon overlijdt zonder dat dit wordt geregistreerd in de GBA

### **2.2 HERKENNEN VAN HEROPNAMEN**

Het herkennen van heropnamen kan langs twee hoofdwegen geschieden:

- a. door het vergelijken van LMR-PAT-ID (patiënt identificatie-code in de LMR)
- b. door het vergelijken van combinaties van waarden op partieel identificerende variabelen

Beide wegen worden hieronder uitgewerkt.

*ad a. Herkennen heropnamen door vergelijken LMR-PAT-ID*

LMR-PAT-ID is de combinatie van ziekenhuisnummer en ziekenhuisspecifiek patiëntnummer. Bij een heropname in hetzelfde ziekenhuis is in principe LMR-PAT-ID gelijk. Het herkennen van heropnamen door LMR-PAT-ID vervalt bij:

- ◆ een heropname in een ander ziekenhuis
- ◆ ander gebruik van LMR-PAT-ID dan gedefinieerd. Bijvoorbeeld indien een patiënt binnen dezelfde instelling onder meerdere patiëntnummers bekend is of indien meerdere patiënten hetzelfde patiëntnummer hebben. Van 23 ziekenhuizen in Nederland is het bekend dat LMR-PAT-ID niet kan worden gebruikt om heropnamen te herkennen. Uiteraard kunnen ook bij de resterende ziekenhuizen problemen aanwezig zijn.
- ◆ een wisseling in het systeem van uitgifte van het (ziekenhuisspecifiek) patiëntnummer. Dit kan het geval zijn de overgang naar een nieuw systeem of bij het samengaan van systemen bij fusies tussen ziekenhuizen.

*ad b. Herkennen heropnamen door combineren van partieel identificerende variabelen*

In de LMR zijn diverse partieel identificerende variabelen aanwezig die kunnen worden gebruikt om opnamen behorende tot hetzelfde individu te herkennen. Potentiële variabelen zijn geslacht, geboortedatum, het 4-cijferige postcodedeel, huisartsencode, betalende instantie, verzekeringswijze, etc. Door het combineren van verschillende variabelen kan een sterk identificerend systeem ontstaan, wat kan worden gebruikt om (kans)uitspraken te doen of twee records tot dezelfde persoon behoren (medical record linkage). Problemen kunnen ontstaan:

- ◆ indien de combinatie van variabelen onvoldoende discriminerend vermogen bezit. Dit betekent dat (te veel) verschillende personen een gelijke combinatie van waarden hebben op de gebruikte koppelv variabelen. Zonder additionele informatie zijn deze personen niet van elkaar te scheiden.
- ◆ indien variabelen binnen een persoon in de tijd van waarde kunnen veranderen, zoals bij postcode, verzekeringswijze en huisartsencode. Bij een te strenge definitie van heropnamen, bijvoorbeeld alle beschikbare koppelv variabelen moeten aan elkaar gelijk zijn, worden heropnamen gemist waarbij een verandering is opgetreden in één van de variabelen.
- ◆ indien dezelfde informatie verschillend wordt vastgelegd. Zo is het systeem van coderen van huisartsen ziekenhuis-specifiek. De code van een zelfde huisarts kan daarmee per ziekenhuis verschillen.
- ◆ ten gevolge van invoerfouten. Net als hierboven is er dan sprake van een wijziging in waarde van een variabele, hetgeen kan leiden tot het missen van een heropname.

In deze haalbaarheidsstudie zijn een aantal cruciale stappen bij het herkennen van heropnamen en overlijden nader onderzocht. Het volgende hoofdstuk geeft een beschrijving van de uitgevoerde activiteiten.

### 3. DOEL PILOT EN UITGEVOERDE ACTIVITEITEN

In de pilot kunnen en worden niet alle problemen onderzocht die samenhangen met het herkennen van heropnamen en overlijden binnen cohort-onderzoek. De pilot-studie concentreerde zich op twee kernvragen:

1. Wat is de kwaliteit van de initiële koppeling tussen LMR en GBA records?
2. Wat is de meerwaarde van de GBA bij het herkennen van heropnamen?

Door specifieke koppelingen daadwerkelijk uit te voeren is inzicht verkregen in de technische problemen en de kwaliteit van de verschillende koppelingen. Waar mogelijk zijn problemen gekwantificeerd en achterliggende oorzaken opgespoord. Een ad hoc LMR-cohort heeft model gestaan voor toekomstige (LMR)-cohorten waarbinnen het patroon van ziekenhuisopnamen en sterfte zal worden geanalyseerd. Alle patiënten ouder dan 1 jaar die in januari 1996 uit een Nederlands ziekenhuis zijn ontslagen (geen dagbehandeling), vormden het begincohort. Voor het herkennen van heropnamen werd uitgegaan van een basisbestand met alle ontslagen ziekenhuispatiënten (incl. dagbehandelingen) in de jaren 1996 en 1997. Alvorens de resultaten te tonen, beschrijven wij in het kort het belang van beide vragen en de specifieke activiteiten die zijn uitgevoerd.

#### 3.1 KWALITEIT INITIËLE KOPPELING TUSSEN LMR EN GBA

De initiële koppeling LMR\*GBA is van cruciaal belang. Deze koppeling heeft de volgende belangrijke functies:

- ◆ via de GBA is het mogelijk om ook het overlijden van personen buiten het ziekenhuis te achterhalen. Bovendien kan langs deze weg de doodsoorzaak worden achterhaald, zowel van personen die binnen als buiten het ziekenhuis zijn overleden.
- ◆ de GBA laat zien welke personen uit het LMR-cohort geen unieke combinatie van waarden hebben op potentiële koppelv variabelen (een niet-unique sleutel). Deze informatie kan worden gebruikt bij het koppelen van LMR-records onderling, bijvoorbeeld door alleen verder te gaan met personen met een unieke sleutel in de GBA.

#### **Uitgevoerde activiteiten**

Bij het terugvinden van leden van het LMR-cohort in de GBA moet gebruik worden gemaakt van combinaties van *partieel* identificerende variabelen die zowel in de LMR als in de GBA aanwezig zijn. Hiertoe behoren de volgende variabelen geslacht, geboortedatum en het 4-cijferige postcodedeel. Het identificerend vermogen van deze

set van variabelen staat daarom centraal. Onder ‘identificerend vermogen’ wordt hierbij verstaan de mate waarin individuen uniek kunnen worden getraceerd middels een 1:1 koppeling, zonder hierbij overigens over naamgegevens te beschikken (anonieme koppeling). Dit heeft geleid tot de volgende activiteiten om meer inzicht te krijgen in de kwaliteit van de initiële LMR\*GBA koppeling:

### **3.1.1 Uniciteit primaire koppelsleutel in GBA**

Hierbij wordt de uniciteit bepaald van de sleutel geslacht, geboortedatum en 4-cijferig postcodedeel in de GBA (heel Nederland), met als deelvragen de variatie in uniciteit afhankelijk van waarden op variabelen uit de sleutel, zoals geslacht, geboortedatum, aantal inwoners van postcodegebied en etniciteit.

### **3.1.2 Uniciteit primaire koppelsleutel in LMR-cohort**

Het bepalen van de uniciteit van dezelfde sleutel binnen een LMR-cohort met speciale aandacht voor het vergelijken van de mate van uniciteit in de GBA met die in het LMR-cohort.

### **3.1.3 GBA koppelstatus van patiënten uit LMR-cohort**

Bij het koppelen van GBA en LMR records op basis van een bepaalde koppelsleutel kunnen zich in totaal negen situaties voordoen, te weten:

		LMR		
		0	1	> 1
GBA	0	*	<i>a</i>	*
	1	*	<i>b</i>	*
	> 1	*	<i>c</i>	*

Voor dit onderzoek zijn er echter maar drie van belang, tezamen aangeduid met de term “GBA-koppelstatus”: LMR record is 'uniek in GBA' (*b*), 'niet-uniek in GBA' (*c*), 'niet aanwezig in GBA' (*a*); de overige mogelijkheden (\*) worden buiten beschouwing gelaten. Van de patiënten die eenmaal uniek in de GBA zijn herkend, kan het overlijden (plus de bijbehorende datum) worden achterhaald. Ook dit vervolgtraject is technisch uitgevoerd. Deze informatie is op twee manieren gebruikt:

### **3.1.4 Patroon van overlijden in LMR-cohort**

Het beschrijven van het patroon van overlijden gedurende twee jaar na ontslag van patiënten uit het LMR januari cohort die uniek koppelden in de GBA. Verschillen in overlijden tussen patiënten met verschillende aandoeningen zijn nagegaan.

### **3.1.5 Verkeerde koppelingen tussen LMR en GBA records**

Discrepancies in patiënten die volgens de LMR in het ziekenhuis zijn overleden maar nog in leven zijn volgens de GBA, wijzen sterk in de richting van onjuiste koppelingen tussen LMR en GBA records.

## 3.2 MEERWAARDE GBA VOOR HERKENNEN HEROPNAMEN

Dit is een vergelijking tussen enerzijds een koppeling LMR\*LMR met alleen initieel een koppeling met de GBA (voor bepalen uniciteit sleutel en het bieden van een entree tot de doodsoorzakenstatistiek) en anderzijds een koppeling LMR\*LMR waarbij de GBA ook tijdens de follow-up wordt gebruikt voor het up-to-date houden van postcodes en het controleren van de initiële uniciteit. De potentiële meerwaarde van het up-to-date houden van de postcode via de GBA dient te worden afgewogen tegen een flinke toename in complexiteit, en dus in tijd en kosten. Hieronder staan de potentiële problemen die ontstaan bij het herkennen van heropnamen zonder verdere tussenkomst van de GBA (de eerst genoemde optie):

- ◆ een initieel uniek gekoppeld persoon in de GBA kan gedurende follow-up niet-uniek worden doordat een persoon van buiten het cohort verhuist waarbij zijn of haar sleutel gelijk wordt aan een persoon in het cohort. Opnamen van dergelijke administratieve wederhelten kunnen leiden tot onjuiste heropnamen (fout-positieve koppelingen).
- ◆ verschillende potentiële koppelv variabelen in de LMR kunnen binnen een lid van het cohort (legitiem) van waarde veranderen zoals postcode, huisartsencode en verzekeringswijze. Hierdoor kunnen heropnamen worden gemist indien de koppeling sterk leunt op een dergelijke variabele. Vooral postcode en huisartsencode zijn belangrijke variabelen in een LMR\*LMR koppeling. Via de GBA is een update mogelijk van de variabele postcode, waarbij tevens de uniciteit van de nieuwe koppelsleutel in de GBA kan worden nagegaan. Dit kan de kwaliteit van de LMR\*LMR koppeling verbeteren, vooral bij het herkennen van heropnamen van cohortleden die verhuizen *en* in een ander ziekenhuis dan de index-opname worden heropgenomen.

### Uitgevoerde activiteiten

Binnen de pilot ontbraken de tijd en middelen om de twee strategieën voor het herkennen van heropnamen (alleen initieel GBA versus GBA ook tijdens follow-up) rechtstreeks met elkaar te vergelijken. De meerwaarde van een LMR\*LMR koppeling met gebruik van de GBA tijdens follow-up is indirect geschat door onderzoek te doen naar de problemen van het herkennen van heropnamen zonder gebruik van de GBA tijdens follow-up:

#### 3.2.1 *Verhuispatroon alle Nederlanders*

Dit betreft een analyse van het verhuispatroon van alle Nederlanders binnen een kalenderjaar naar leeftijd, geslacht en het wel of niet veranderen van het 4-cijferige deel van de postcode (één van de aanstichters van problemen).

#### 3.2.2 *Verhuispatroon LMR-cohort*

Dit is een vergelijkbare analyse als hierboven, maar nu wordt het verhuispatroon van een algemeen LMR cohort (zieke mensen) onderzocht. Het cohort wordt zowel in zijn geheel als naar enkele hoofdgroepen van diagnoses geanalyseerd.

### **3.2.3 Verlies unieke GBA koppelstatus in de tijd**

In de tijd kunnen (LMR) patiënten die eerst uniek waren in de GBA niet-uniek worden op de primaire koppelsleutel, bijvoorbeeld door het in de buurt komen wonen van iemand met hetzelfde geslacht en dezelfde geboortedatum. Wij hebben geteld hoeveel GBA-unieke mensen uit het LMR-cohort na één kalenderjaar niet meer uniek waren.

### **3.2.4 Heropnamefrequentie in relatie tot verhuizen**

De meerwaarde van het gebruik van de GBA tijdens de follow-up is vooral gelegen in het beter kunnen herkennen van heropnamen van patiënten die verhuizen, in het bijzonder wanneer de daarop volgende heropnamen plaatsvinden in een ander ziekenhuis dan de index-opname. 'Index-opname' is hier gedefinieerd als de eerste opname (excl. dagopnamen) van een patiënt in het begincohort (januari 1996); het gaat hier *niet* om eerste opnamen voor een bepaalde ziekte. Door het vergelijken van de frequentie van heropnamen (inclusief dagopnamen) van niet-verhuizers en verhuizers wordt een schatting gemaakt van de potentiële meerwaarde van het gebruik van de GBA tijdens follow-up.

## 4. RESULTATEN MET KORTE BESPREKING

### 4.1 RESULTATEN KWALITEIT INITIËLE KOPPELING TUSSEN LMR EN GBA

#### 4.1.1 Unicité primaire koppelsleutel in GBA

De GBA bevat demografische gegevens, waaronder adresgegevens, van alle in Nederland ingeschreven bewoners. De structuurtelling van 1-1-1996 is gebruikt om na te gaan hoeveel personen in Nederland een unieke combinatie van waarden hadden op de variabelen geslacht, geboortedatum (dag/maand/jaar) en het 4-cijferige deel van de postcode (primaire koppelsleutel). De resultaten staan samengevat in tabel 1.

**Tabel 1.** Unicité van de primaire koppelsleutel (geslacht, geboortedag, -maand, -jaar en cijferdeel postcode) in de Gemeentelijke Basis Administratie. Peildatum 1-1-1996.

Unicité sleutel	Aantal codes <sup>1</sup>	Aantal personen	% personen
Eenling	13.297.720	13.297.720	85,8
Administratieve tweeling <sup>2</sup>	963.839	1.927.678	12,4
Administratieve drieling <sup>2</sup>	73.153	219.459	1,4
Administratieve vierling of hoger <sup>2</sup>	9.946	49.032	0,3
Totaal	14.344.658	15.493.889	100

<sup>1</sup> het aantal verschillende combinaties van waarden op de koppelsleutel in de GBA

<sup>2</sup> administratieve n-ling: n personen met dezelfde combinatie van waarden op de koppelsleutel

Deze gegevens laten zien dat 86% van alle Nederlanders een unieke combinatie van waarden (sleutel) heeft op bovengenoemde variabelen. Het merendeel (88%) van de niet-unieke personen heeft slechts één ander persoon met dezelfde code; zij worden hier aangeduid als administratieve tweelingen.

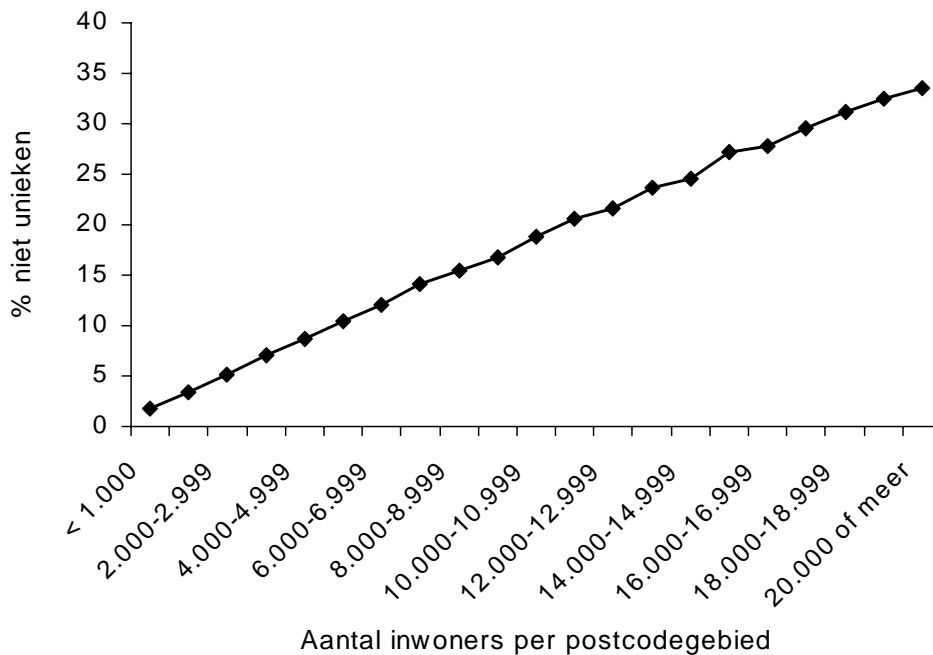
Er is onderzocht of het percentage personen met een unieke sleutel afhangt van de waarden van variabelen uit de sleutel. De relatie tussen de leeftijd en het percentage unieke personen is weergegeven in tabel 2, de relatie tussen de grootte van het postcodegebied en het percentage niet-unieke personen staat afgebeeld in figuur 3.

**Tabel 2.** Unicité van de primaire koppelsleutel (geslacht, geboortedag, -maand, -jaar en cijferdeel postcode) naar leeftijd in GBA. Peildatum 1-1-1996.

Leeftijdsklasse (in jaren)	% personen met een unieke sleutel	Bevolkingsgrootte <sup>1</sup>
0-4	85,0	19,6
5-24	85,4	19,4
25-44	82,6	25,1
45-54	85,7	21,0
55-64	89,6	14,7
65-74	91,0	11,8
75-84	93,5	6,7
85+	96,7	1,4
Alle leeftijden	85,8	15,5

<sup>1</sup> gemiddeld aantal inwoners (x 10.000) per leeftijdsjaar in de betreffende leeftijdsklasse

Het percentage unieke personen tussen 0 en 4 jaar ligt op 85,0%, daalt daarna licht tot 82,6% voor personen tussen de 25 en 44 jaar oud om vervolgens weer toe te nemen tot 91,0% voor personen tussen de 65-74 jaar en tot 96,7% onder mensen ouder dan 85 jaar. Dit patroon volgt ruwweg de vorm van de bevolkingspiramide in Nederland, waarbij de leeftijdsgroepen 25 tot 40 jaar de grootste groepen vormen. Er geldt hoe meer personen aanwezig, hoe groter de (toevals)kans dat verschillende personen toch een zelfde sleutel hebben.



**Figuur 3.** Percentage personen met een niet-unieke combinatie van waarden op geslacht en geboortedatum per 4-cijferig postcodegebied als functie van het aantal inwoners van dat postcodegebied. Cijfers voor heel Nederland per 1-1-1996.

Door de forse verschillen in het aantal inwoners per postcodegebied (4-cijfer deel wel te verstaan) zien we hetzelfde fenomeen ook hier terug. Bij een groter aantal personen met dezelfde postcode stijgt de kans dat door toeval geslacht en geboortedatum ook gelijk zijn. Aannemelijk is gemaakt (zie bijlage) dat het percentage niet-unieke personen lineair toeneemt met het aantal inwoners binnen een bepaald postcodegebied. Dus: als het aantal inwoners verdrievoudigt, neemt ook het percentage niet-unieke personen met een factor 3 toe; dit houdt in dat het aantal niet-unieke personen toeneemt met een factor 9 ( $3 \times 3$ ). Dit zien we terug in figuur 3. Bij een postcode gebied met 4.000-4.999 inwoners ligt het percentage niet-unieke personen rond de 7%, in een postcode gebied met 13.000-13.999 is dit opgelopen tot 22%, ruwweg een verdrievoudiging. Ter oriëntatie, ongeveer 7,3 miljoen Nederlanders wonen in een 4-cijferig postcodegebied met 8.000 of meer inwoners, hiervan woont  $\pm 1$  miljoen in een postcodegebied van 14.000 of meer. Postcodegebieden met veel inwoners bevinden zich voornamelijk in de grote steden.



Als specifiek probleem is nog gekeken naar de geboortedatums van 'allochtonen' in Nederland. Bij allochtonen zijn bepaalde geboortedatums vaker ingevuld, zoals 1 januari en 1 juli, wanneer de exacte datum onbekend was. Ook dit levert bij het koppelen van gegevens problemen op omdat de (toevals)kans op een identieke geboortedatum hierdoor groter wordt. In de GBA was deze voorkeur voor bepaalde datums onder allochtonen goed waarneembaar. In totaal waren 2,66 miljoen allochtonen ingeschreven, waarbij als (ruime) definitie werd gehanteerd dat het geboorteland van de persoon zelf of tenminste één der ouders niet-Nederlands moest zijn. Indien alle geboortedatums even vaak zouden voorkomen, zouden er per geboortedag gemiddeld 7.291 allochtonen ingeschreven staan. Extreme uitschieters waren echter 1 januari met 25.922 personen en 1 juli met 15.551. Sowieso kwam iedere eerste dag van de maand vaker voor ( $\pm 9.400$ ) en, in iets mindere mate, de 'mooie' getallen 10, 15 en 20 van iedere maand.

#### 4.1.2 Uniciteit primaire koppelsleutel in LMR-cohort

Het LMR januari-cohort bestaat uit alle klinische opnamen van personen ouder dan 1 jaar die in januari 1996 uit het ziekenhuis zijn ontslagen. Dit cohort bevat 132.257 opnamen. Binnen het totaal van opnamen is gekeken naar de verdeling van combinaties van waarden op geslacht, geboortedatum en 4-cijferig postcodedeel (zie tabel 3). Deze analyse komt overeen met die van de totale GBA (tabel 1).

**Tabel 3.** Uniciteit van de primaire koppelsleutel (geslacht, geboortedag, -maand, -jaar en cijferdeel postcode) in het LMR januari-cohort van 1996 (n=132.257 records).

Uniciteit sleutel	Aantal codes <sup>1</sup>	Aantal records <sup>3</sup>	% records
Eenling	117.828	117.828	89,1
Administratieve tweeling <sup>2</sup>	6.010	12.020	9,1
Administratieve drieling <sup>2</sup>	686	2.058	1,6
Administratieve vierling of hoger <sup>2</sup>	74	351	0,3
Totaal	124.598	132.257	100

<sup>1</sup> het aantal verschillende combinaties van waarden op de koppelsleutel in het LMR-cohort

<sup>2</sup> administratieve n-ling: n opnamen met dezelfde combinatie van waarden op de koppelsleutel

<sup>3</sup> aantal opnamen in LMR-cohort

Het percentage unieke records in het LMR-cohort ligt op 89%. Dit percentage is niet (direct) vergelijkbaar met het percentage uniekelingen in de GBA. De niet-unieke records in het LMR-cohort zijn niet allemaal administratieve meerlingen in de GBA. Heropnamen zijn een belangrijke reden voor het ontstaan van niet-unieke records in het LMR cohort. Gezien de korte tijdsperiode van één maand hebben index-opname en heropname hoogst waarschijnlijk dezelfde sleutel.

Om het percentage niet-uniek zijn in het LMR cohort te vergelijken met dat in de GBA dienen heropnamen zo goed mogelijk te worden verwijderd. Wij hebben hier toe die ziekenhuizen geselecteerd die op de gedefinieerde wijze gebruik maken van LMR-PAT-ID (n=102.108 records). Vervolgens zijn heropnamen, gedefinieerd als opnamen waarbij het LMR-PAT-ID gelijk is aan een eerdere opname uit dezelfde maand verwijderd (n=3.695). Opnieuw is de uniciteit van de primaire sleutel onder-

zocht. Door heropnamen zoveel mogelijk te verwijderen geven deze analyses meer inzicht in het voorkomen van niet-unieke *personen* in het LMR cohort in plaats van niet-unieke *records*.

**Tabel 4.** Uniceit van de primaire koppelsleutel (geslacht, geboortedag, -maand, -jaar en cijferdeel postcode) in personen van een subset van ziekenhuizen uit LMR-cohort (n=102.108 opnamen). Heropnamen, gedefinieerd als LMR-PAT-ID gelijk aan een eerdere opname, zijn verwijderd (n=3.695).

Uniceit sleutel	Aantal codes	Aantal personen	% personen
Eenling	94.520	94.520	96,0
Administratieve tweeling	1.889	3.778	3,8
Administratieve drieling	33	99	0,1
Administratieve vierling	4	16	0,0
Totaal	96.446	98.413	100

De schatting van het percentage unieke personen in het bewerkte LMR-cohort komt nu uit op 96%. Ook dit percentage is waarschijnlijk te laag omdat niet alle heropnamen zijn herkend en verwijderd. Vooral heropnamen in een ander ziekenhuis dan de index-opname zijn nog aanwezig. Deze heropnamen hebben een gelijke sleutel, maar een verschillende LMR-PAT-ID.

#### *Vergelijken uniceit in GBA met uniceit in LMR*

In vergelijking met GBA is het percentage niet-unieke personen in het bewerkte LMR-cohort een factor 3,5 lager (14% vs. 4%). Uiteraard valt het percentage administratieve meerlingen in een steekproef (bewerkt LMR-cohort) lager uit dan in de gehele populatie (GBA). Als van een administratieve tweeling in de GBA één persoon wordt opgenomen in het ziekenhuis levert dit in het LMR cohort een (schijnbaar) uniek persoon op. Pas wanneer beiden worden opgenomen, ontstaat ook in de LMR een administratieve tweeling. Onder verwaarlozing van het feit dat de GBA structuurtelling een momentopname is en de LMR een maandperiode beslaat, kan op theoretische gronden worden bewezen (zie appendix) dat het percentage administratieve meerlingen over een groot traject lineair toeneemt met de grootte van de steekproef.

In onze situatie is de steekproef  $\pm 150$  keer zo klein als de populatie (100.000 vs. 15 miljoen). Het percentage administratieve meerlingen zou daarmee rond de 0,1% moeten liggen ( $14,2\%/150$ ). In werkelijkheid ligt het percentage beduidend hoger. Er zijn twee belangrijke verklaringen voor dit verschil. Ten eerste, niet alle heropnamen zijn geïdentificeerd waardoor het aantal administratieve meerlingen te hoog wordt ingeschat (zie hierboven). Ten tweede, LMR-patiënten zijn geen afspiegeling van de algemene Nederlandse bevolking (geen aselechte steekproef): personen van middelbare en hoge leeftijd zijn oververtegenwoordigd. De kans op het ontstaan van administratieve meerlingen is hierdoor groter.

### 4.1.3 GBA koppelstatus van patiënten uit LMR-cohort

Gegeven een bepaalde combinatie van waarden op koppelingsvariabelen (sleutel) kunnen zich 4 situaties voordoen bij een koppeling tussen LMR records en GBA records:

- ◆ één opname uit LMR koppelt met één persoon uit GBA
- ◆ één opname uit LMR koppelt met meerdere personen uit GBA
- ◆ meerdere opnamen uit LMR koppelen met één persoon uit GBA
- ◆ één opname uit LMR koppelt met geen één persoon uit GBA

De term koppelt heeft hier de betekenis van het hebben van een zelfde samengestelde code (=volledige deterministische koppeling), dat wil zeggen alle koppelingsvariabelen stemmen overeen. Bij de koppeling met de GBA zijn wij uitgegaan van alle verschillende codes die voorkwamen in het LMR-cohort (n=124.598, zie tabel 3), dit kan beschouwd worden als een cohort van klinische patiënten ontslagen in de maand januari 1996 waarvan de heropnamen in dezelfde maand (gedefinieerd als een latere opname met dezelfde waarden op de koppelsleutel) zijn verwijderd. De codes van dit patiëntencohort hebben wij proberen terug te vinden in de GBA structuurtelling van 1 januari 1996. In tabel 5 staat de GBA koppelstatus van deze codes ('patiënten') vermeld. Het percentage GBA uniekelingen in dit LMR-cohort ligt op 84,2%.

**Tabel 5.** GBA koppelstatus van alle verschillende codes in het LMR januari-cohort (n=124.598).

GBA koppelstatus <sup>1</sup>	Aantal codes	%
Uniek in GBA	104.889	84,2
Niet-uniek in GBA	13.591	10,9
waarvan 2x voor	12.268	
waarvan 3x voor	1.176	
waarvan 4x voor	121	
waarvan 5 of meer keer	26	
Komt niet voor in GBA	6.118	4,9
Totaal	124.598	100

<sup>1</sup> frequentie van voorkomen van de LMR-codes in de structuurtelling van 1 januari 1996

Een aantal LMR-codes (4,9%) werd niet in de GBA teruggevonden. Deels zijn dit buitenlanders die tijdens hun vakantie of bezoek aan Nederland in het ziekenhuis belanden. De rest heeft waarschijnlijk te maken met personen die bewust niet zijn ingeschreven (illegalen); personen die bewust verkeerde gegevens aanleveren; administratieve fouten (verschillen) in het vastleggen van koppelingsvariabelen in LMR en/of GBA. Dit laatste probleem zal zeker bij de variabele postcode optreden door verschillen in het vastleggen van moeilijke postcodes (bijvoorbeeld woonboten, afgelegde huizen, nieuwe wijken), het wonen op tijdelijke adressen, door het gebruik van verouderde postcodes in GBA of LMR, of door tussentijdse verhuizingen, hoewel de kans hierop geminimaliseerd is door een januaricohort uit de LMR te nemen en deze te vergelijken met de GBA per 1 januari.

#### 4.1.4 Patroon van overlijden in LMR-cohort

Wanneer patiënten eenmaal zijn geïdentificeerd in de GBA is het mogelijk om langs deze weg de sterfte te achterhalen. Wij hebben voor alle patiënten die uniek koppelden in de GBA (de unieke codes uit tabel 5, n=104.889) de sterfte na ontslag in de jaren 1996 en 1997 geanalyseerd. Voor de goede orde, het gaat hierbij zowel om de sterfte in als buiten het ziekenhuis. Deze totale sterfte, uitgezet naar leeftijd, staat in tabel 6. De tabel toont overduidelijk aan dat de sterftetekansen van een willekeurige groep ontslagen patiënten groot zijn, vooral op hogere leeftijd.

**Tabel 6.** Sterfte volgens de GBA onder patiënten van het LMR-cohort die uniek koppelden binnen de GBA (n=104.889). Cumulatieve sterfte (in %) ½, 1 en 2 jaar na ontslag uit het ziekenhuis.

Leeftijd in januari-cohort	# patiënten in januari-cohort	% sterfte in eerste ½ jaar	% sterfte in eerste jaar	% sterfte in eerste 2 jaar
0-14	5.305	0,7	1,0	1,4
15-34	18.182	0,5	0,8	1,1
35-54	23.847	3,0	4,3	5,7
55-74	34.176	9,9	13,7	18,7
75+	23.379	21,2	27,6	37,5
Alle leeftijden	104.889	8,8	11,8	16,0

Wij hebben ook gekeken naar de invloed van de hoofddiagnose bij de index-opname op de sterfte (zie tabel 7). De indeling van patiënten kwam tot stand op basis van de hoofddiagnose van de index opname in januari 1996. Vooral de sombere prognose van patiënten met een maligniteit en hartfalen vallen op.

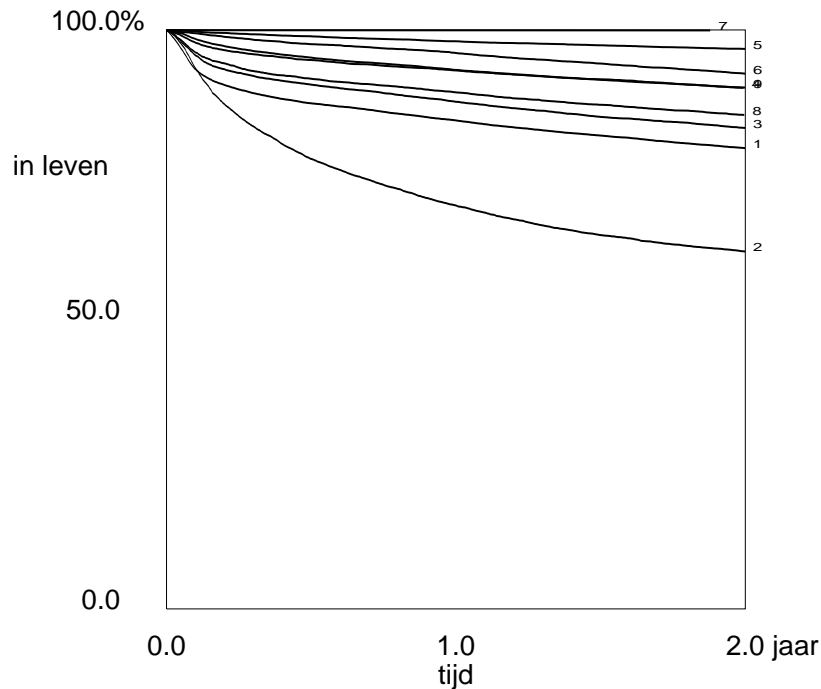
**Tabel 7.** De cumulatieve sterfte volgens de GBA 1 en 2 jaar na ontslag voor verschillende groepen patiënten uit het 'unieke' LMR-januari-cohort (n=104.889).

Hoofddiagnose bij index opname	# patiënten in januari-cohort	% sterfte in eerste jaar	% sterfte in eerste 2 jaar
Alle hart- en vaatziekten	18.159	17,9	23,4
AMI	1.975	23,6	28,2
CVA	2.219	31,9	37,3
HF	2.121	43,4	56,6
PAV	2.282	14,4	20,0
Maligniteiten	9.636	35,2	44,4
Luchtwegen	8.610	14,7	20,1
Spijsvertering	10.157	8,3	11,9
Bewegingsstelsel	10.246	2,3	3,8
Overige patiënten	48.081	7,0	10,3
Alle patiënten	104.889	11,8	16,0

Afkortingen en ICD-9 codes: Hart- en vaatziekten 390-458; AMI = acuut hartinfarct 410; CVA = cerebrovasculaire aandoeningen 430-438; HF = hartfalen 428; PAV = perifere arterieel vaatlijden 440-448; Maligniteiten 140-239; Luchtwegen 460-519; Spijsvertering 520-579; Bewegingsstelsel 710-739

Doordat de GBA niet alleen registreert of iemand is overleden maar ook het bijbehorende tijdstip, is het mogelijk een overlevingscurve te construeren. In figuur 4 is de

totale sterfte na ontslag weergegeven voor verschillende groepen van patiënten. Zoals verwacht bestaan er aanzienlijk verschillen in overleving. De sterfte is het hoogst onder patiënten met een maligniteit en het laagst bij vrouwen die wegens hun zwangerschap waren opgenomen (bijna nihil). Na patiënten met een maligniteit hebben patiënten met een hart- of vaatziekte de minst gunstige prognose. De totale sterfte na twee jaar bedraagt 23%.



**Figuur 4.** Overleving van verschillende patiëntengroepen uit het 'unieke' LMR-cohort tot twee jaar na ontslag uit het ziekenhuis (1=hart- en vaatziekten; 2=Maligniteiten; 3=luchtwegen; 4=spijsvertering; 5=bewegingsstelsel; 6=zenuwstelsel; 7=zwangerschap; 8=symptomen; 9=restgroep).

Bovenstaande analyses tonen aan dat wanneer personen eenmaal uniek zijn herkend in de GBA het patroon van sterfte (inclusief tijdstip) kan worden geanalyseerd, ongeacht de plaats van overlijden (thuis, ziekenhuis, verpleeghuis). Doordat er een directe koppeling mogelijk is tussen GBA en de doodsoorzakenstatistiek is ook een analyse naar doodsoorzaken mogelijk. Dit traject is binnen deze pilot niet uitgevoerd.

#### **4.1.5 Verkeerde koppelingen tussen LMR en GBA records**

Ondanks het feit dat een patiënt uit het LMR bestand uniek koppelde met een persoon uit de GBA, kan het toch om verschillende personen gaan. Een invoerfout in bijvoorbeeld de postcode bij een patiënt in de LMR, kan toevallig leiden tot een unieke koppeling met een (verkeerde) persoon in de GBA. We hebben naar aanwijzingen hiervoor gezocht door te kijken of alle GBA-unieke patiënten die in het ziekenhuis zijn overleden tijdens de index-opname ook in de GBA zijn overleden. Als dit niet het geval was dan is of het overlijden foutief vastgelegd in de LMR, of het overlijden gemist in de GBA, of er is sprake van een invoerfout in GBA of LMR met als gevolg

een foutieve koppeling. Vermoedelijk is vooral de laatste reden van belang omdat de eerste twee oorzaken zeer onwaarschijnlijk zijn. Van alle LMR-patiënten die uniek koppelden in de GBA (n=104.889) overleden 4.178 patiënten in het ziekenhuis tijdens de index-opname. Van deze groep waren 4.143 personen ook overleden volgens de GBA. Bij 35 patiënten was er dus sprake van een discrepantie (0,8%).

## 4.2 RESULTATEN MEERWAARDE GBA VOOR HERKENNEN HEROPNAMEN

Dit is de eerste activiteit waarin het herkennen van heropnamen centraal staat. Voor het koppelen van LMR-records onderling zijn postcode en/of huisartsencode belangrijke variabelen. Een nadeel van deze variabelen is het kunnen veranderen van waarde bij bijvoorbeeld verhuizen. Dit maakt vooral het herkennen van heropnamen van patiënten die èn verhuizen èn worden opgenomen in een ander ziekenhuis lastig. Als eerste is daarom gekeken naar het verhuisgedrag van alle Nederlanders (paragraaf 4.2.1) en van opgenomen patiënten in het bijzonder (paragraaf 4.2.2).

### 4.2.1 Verhuispatroon alle Nederlanders

Het verhuisgedrag van alle Nederlands ingezetenen (hierna kortweg Nederlanders genoemd) in het jaar 1996 is geanalyseerd met behulp van de stroomgegevens uit de GBA. De resultaten aangaande 'verre' verhuizingen, dat wil zeggen verhuizingen waarbij het 4-cijferige van de postcode verandert, staan samengevat in tabel 8. De gemiddelde jaarlijkse kans om te verhuizen bedraagt 7,5%. Jongeren tussen de 15-34 jaar verhuizen het meest (jaarlijkse kans van 14%). Ook de leeftijdsklasse van 75 jaar en ouder heeft weer een wat hogere verhuiskans dan de voorafgaande leeftijdsklasse. Waarschijnlijk hangt dit samen met het kleiner, dan wel aangepast gaan wonen.

**Tabel 8.** Aantal Nederlanders dat 1 of meer keer verhuisde in 1996 naar leeftijd. Alleen 'verre' verhuizingen waarbij het 4-cijferige deel van de postcode veranderde, zijn geteld.

Leeftijd op 1-1-1996	# personen met 1 of meer 'verre' verhuizingen	Totaal aantal Nederlanders	verhuis % per jaar
0-14	168.512	2.861.353	5,9
15-34	642.411	4.562.427	14,1
35-54	233.080	4.577.714	5,1
55-74	83.316	2.665.124	3,1
75+	44.918	900.489	5,0
Alle leeftijden	1.172.237	15.567.107	7,5

### 4.2.2 Verhuispatroon LMR-cohort

Het is belangrijk om te weten of een groep 'zieke' mensen, zoals een LMR-cohort, een vergelijkbaar verhuispatroon heeft. Het is zowel denkbaar dat zij minder vaak verhuizen (te vermoeiend, uit financiële overwegingen) als dat zij vaker verhuizen (kleiner huis, minder trappen, dichterbij zorg). In tabel 9 staat het verhuispatroon beschreven van patiënten uit het LMR-cohort, en wel dat deel dat uniek koppelde in

de GBA (n=104.889). Van deze uniek gekoppelde personen kunnen de verhuizingen worden geanalyseerd via de GBA stroomgegevens.

Het verhuispatroon van een willekeurig LMR-cohort komt in grote lijnen overeen met die van de algemene Nederlandse bevolking (vergelijk kolom 4 van tabel 9 met de laatste kolom van tabel 8). Wel is de grote verhuiskans van de groep 15-34 jarigen in de algemene bevolking binnen het LMR-cohort minder uitgesproken. Indien we het verhuizen analyseren in de tijd dan valt op dat de kans hierop vrijwel constant is in de eerste twee jaar na ontslag.

**Tabel 9.** Percentage patiënten uit het 'unieke' LMR-cohort (n=104.889) dat 1 of meer keer is verhuisd op een ½, 1 en 2 jaar na ontslag uit het ziekenhuis. Alleen 'verre' verhuizingen waarbij het 4-cijferige deel van de postcode veranderde, zijn geteld.

Leeftijd in januari-cohort	# patiënten in januari-cohort	% verhuisd in eerste ½ jaar	% verhuisd in eerste jaar	% verhuisd in eerste 2 jaar
0-14	5.305	2,4	6,2	11,7
15-34	18.182	4,2	10,8	21,2
35-54	23.847	1,9	4,5	8,8
55-74	34.176	1,3	3,1	6,2
75+	23.379	3,6	6,4	10,1
Alle leeftijden	104.889	2,5	5,7	10,8

Bij het vergelijken van de verhuispercentages van LMR-patiënten met die van heel Nederland is geen rekening gehouden met tussentijdse sterfte. Het aanzienlijke verschil in sterfte tussen LMR patiënten en de algemene Nederlandse bevolking maakt dat de verhuiskans van nog in leven zijnde LMR-patiënten hoger ligt dan de percentages in tabel 9. Door gebruik te maken van de sterfte in het LMR-cohort (zie hiervoor) kan het belang van tussentijds overlijden op de overall verhuiskans worden nagegaan. De verhuiskans in het eerste jaar van de gehele groep gaat van 5,7% naar 6,4% indien als noemer het aantal patiënten wordt genomen dat na 1 jaar nog in leven is. Op dezelfde manier loopt de verhuiskans voor groep ouder dan 75 jaar op van 6,4 tot 8,9%.

Ook is de invloed van de primaire aandoening op het patroon van verhuizen onderzocht. Voor verschillende ziektebeelden (indeling op basis hoofddiagnose bij indexopname) is geteld hoeveel patiënten één of meer keer verhuisden in 1996 of in 1997. Verhuizen is ook hier weer gedefinieerd als het veranderen van het 4-cijfer deel van de postcode. De resultaten staan in tabel 10.

Tussen ziektebeelden verschilt het patroon van verhuizen aanzienlijk, in het uiterste geval een factor twee. Het lage percentage verhuizen onder patiënten met een maligniteit hangt gedeeltelijk samen met de hogere sterfte in deze groep. Het hoge percentage verhuizen onder patiënten met een cerebrovasculaire aandoening (CVA) is mogelijk het gevolg van het aangepast gaan wonen vanwege de restverschijnselen van het CVA.

**Tabel 10.** Aantal patiënten uit het 'unieke' LMR januari cohort (n=104.889) dat minimaal één keer verhuisde met een verandering in 4-cijferdeel postcode in de 2 jaren na ontslag. Indeling naar hoofddiagnosen bij de index-opname.

Hoofddiagnose bij Index opname	# patiënten in januari-cohort	# patiënten met minimaal 1 'verre' verhuizing	%
Alle hart- en vaatziekten	18.159	1.482	8,2
AMI	1.975	96	4,9
CVA	2.219	281	12,7
HF	2.121	182	8,6
PAV	2.282	176	7,7
Maligniteiten	9.636	542	5,6
Luchtwegen	8.610	1.112	12,9
Spijsvertering	10.157	1.032	10,2
Bewegingsstelsel	10.246	1.044	10,2
Overige patiënten	48.081	6.163	12,8
Alle patiënten	104.889	11.375	10,8

AMI = acuut hartinfarct; CVA = cerebrovasculaire aandoeningen; HF = hartfalen; PAV = perifere arterieel vaatlijden. Voor ICD-codes zie tabel 7

### 4.2.3 Verlies unieke GBA koppelstatus in de tijd

Wij hebben geanalyseerd wat de kans is dat mensen die eerst uniek waren in de GBA later niet-uniek worden. Dit niet-uniek worden kan op twee manieren. De persoon in kwestie verhuist zelf naar een postcodegebied waar al iemand woont met een gelijk geslacht en een zelfde geboortedatum. Of iemand wordt niet-uniek doordat een ander komt te wonen in zijn of haar postcodegebied met hetzelfde geslacht en geboortedatum. Het verlies aan uniciteit is onderzocht door van alle LMR-patiënten die uniek waren in de GBA structuurtelling van 1-1-1996 opnieuw hun status te bepalen in de structuurtelling van 1-1-1997. De resultaten staan in tabel 11.

**Tabel 11.** GBA koppelstatus op 1-1-1997 van alle LMR-patiënten die uniek waren in de GBA op 1-1-1996 (n=104.889).

GBA koppelstatus op 1-1-1997	Aantal personen	%
Uniek gebleven in GBA	91.762	87,5
Niet-uniek geworden in GBA	899	0,9
Komt niet meer voor in GBA	12.228	11,7
Alle	104.889	100

Minder dan 1% van alle unieke op 1-1-1996 had zijn/haar unieke status verloren op 1-1-1997 en was deel gaan uitmaken van een administratieve meerling. Bijna 12% (12.228) van de LMR-patiënten kwam niet meer voor in de GBA. De belangrijkste oorzaak hiervan is het overlijden van LMR-patiënten. Bij ruim 12.000 LMR-patiënten is het niet meer voorkomen in de GBA het gevolg van overlijden. Andere potentiële oorzaken van het niet meer voorkomen in de GBA zijn emigratie of het niet meer woonachtig zijn op het oude adres zonder dat een nieuw adres bekend is. Dit aantal 'lost to follow-up' lag rond de 200 (0,2% per jaar).



#### 4.2.4 Heropnamefrequentie in relatie tot verhuizen

Van alle patiënten uit het LMR-cohort die met één persoon in de GBA koppelden, is getracht hun heropnamen te achterhalen in de resterende maanden van 1996 en in heel 1997. Voor het herkennen van heropnamen zijn verschillende definities gebruikt die hieronder zijn weergegeven. Bij het herkennen van heropnamen is uitsluitend gebruik gemaakt van informatie (variabelen) uit de LMR. In het bijzonder is in deze pilot geen gebruik gemaakt van informatie uit de GBA om de koppeling binnen de LMR te verbeteren, bijvoorbeeld door de postcode aan te passen wanneer patiënten verhuizen.

**Tabel 12.** Definitie en het aantal heropnamen in 1996 en 1997 van 104.889 patiënten uit het januari 1996 die initieel uniek koppelden in de GBA.

Definitie heropname (code)	Aantal heropnamen
LMR-PAT-ID gelijk + primaire koppelsleutel <sup>1</sup> gelijk (A)	63.970
LMR-PAT-ID gelijk + primaire koppelsleutel verschillend (B)	2.229
alleen postcode verschillend (B1)	2.204
geboortedatum en/of geslacht verschillend (B2)	25
LMR-PAT-ID verschillend + primaire koppelsleutel gelijk (C)	23.567
zelfde ziekenhuis als index + HA-code gelijk (C1)	5.867
zelfde ziekenhuis als index + HA-code verschillend + verzekering en betalende instantie gelijk (C2)	516
zelfde ziekenhuis als index + HA-code en verzekering en betalende instantie verschillend (C3)	377
ander ziekenhuis dan index + verzekering en betalende instantie gelijk (C4)	6.855
overige heropnamen in andere ziekenhuizen (C5)	9.952
LMR-PAT-ID en primaire koppelsleutel verschillend maar wel (D)	42.730
opname in zelfde ziekenhuis + geslacht en geb. datum en HA en verzekering en betalende instantie gelijk maar postcode verschillend (D1)	220
opname in zelfde ziekenhuis + geslacht en geb. datum en HA gelijk + verzekering en/of betalende instantie verschillend + postcode verschillend (D2)	1.299
opname in ander ziekenhuis + geslacht en geb. datum en verzekering en betalende instantie gelijk + postcode verschillend (D3)	41.211
Alle heropnamen zoals hierboven beschreven	132.496

<sup>1</sup> primaire koppelsleutel is combinatie van waarden op de variabelen geslacht, geboortedatum en 4-cijferig deel postcode ten tijde van index opname  
HA-code = huisartsencode

Ondanks de veelheid aan definities zullen niet alle heropnamen herkend zijn. Heropnamen waarbij het aantal verschillen in vergelijking met de index-opname te groot is, bijvoorbeeld door verhuizen in combinatie met een andere huisarts, een andere verzekeringswijze en een heropname in ander ziekenhuis blijven onopgemerkt. Ook de aanwezigheid van meerdere administratieve fouten leidt tot het missen van heropnamen.

Niet alle categorieën in tabel 12 zijn even betrouwbaar, in het bijzonder geldt dit voor categorie D3. Het aantal variabelen dat dient overeen te stemmen in deze categorie is gering (weinig discriminerend vermogen). Dit heeft als consequentie dat heropnamen van niet-cohortdeelnemers wel aan cohortdeelnemers worden toegeschreven. Toch is een deel van categorie D3 wel van belang: het bevat een deel van de heropnamen van patiënten die én zijn verhuisd én in een ander ziekenhuis zijn heropgenomen. Deze heropnamen zijn op basis van LMR informatie alleen, niet te herkennen. De meerwaarde van het gebruik van de GBA in de follow-up is dan ook gelegen in het beter herkennen (onderscheiden) van deze heropnamen. Ondanks dat een directe vergelijking (zonder versus met GBA in follow-up) niet is uitgevoerd, hebben wij toch een schatting gemaakt van de (potentiële) meerwaarde van de GBA tijdens de follow-up.

#### *Heropnamen en verhuizen: potentiële meerwaarde van de GBA tijdens follow-up*

Langs indirecte weg hebben wij getracht te schatten hoe groot het aantal heropnamen is van mensen die én verhuisd zijn én in een ander ziekenhuis zijn opgenomen. Deze schatting is gebaseerd op de volgende werkwijze en aannames:

1. alle patiënten die uniek koppelden in de GBA deelden wij in als verhuizers (één of meer keer verhuisd in 1996 of 1997 waarbij cijfercombinatie van postcode veranderde) dan wel niet-verhuizers. Dit indelen geschiedde aan de hand van GBA verhuisberichten.
2. alle heropnamen van niet-verhuizers kunnen worden achterhaald. Voor heropnamen in hetzelfde ziekenhuis als de index-opname gebruikten wij de definities A+C1+C2+C3 (C1 tot en met C3 voor heropnamen waarbij LMR-PAT-ID niet correct is toegepast). Heropnamen in een ander ziekenhuis werden herkend aan de hand van de (initieel) unieke primaire koppelsleutel: definities C4+C5.

Voor verhuizers ligt de zaak complexer:

3. alle heropnamen (uit index-ziekenhuis of elders) van verhuizers die plaatsvonden voor het daadwerkelijke tijdstip van verhuizen, kunnen worden herkend (zelfde definities als in 2).
4. alle heropnamen na het tijdstip van verhuizen, maar naar hetzelfde ziekenhuis als de indexopname, kunnen worden herkend via LMR-PAT-ID of door de combinatie van gelijk geslacht, geboortedatum en huisartsencode (vergelijkbaar identificerend vermogen als postcode). Dit zijn heropnamen die vallen onder de definities B1, D1 en D2.
5. voor het herkennen van heropnamen na verhuizen in een ander ziekenhuis bestaan geen betrouwbare methoden in de LMR; deze worden gemist.

Tabel 13 geeft een overzicht van het aantal heropnamen van niet-verhuizers en verhuizers zoals die voortkomen uit de verschillende aannames.

**Tabel 13.** Overzicht van het aantal heropnamen voor patiënten die al dan niet zijn verhuisd tijdens de follow-up. Aantal heropnamen zijn geteld volgens de aannames 2 tot en met 5.

Type heropname	Niet verhuizers (n=93.514)	Verhuizers (n=11.375)
Heropnamen in index ziekenhuis voor verhuizen	66.282	4.448
Heropnamen in ander ziekenhuis voor verhuizen	15.635	1.172
Heropnamen in index ziekenhuis na verhuizen	0 <sup>1</sup>	2.227
Heropnamen in ander ziekenhuis na verhuizen	0 <sup>1</sup>	?
Totaal aantal heropnamen	81.917	7.847+?

<sup>1</sup> per definitie

Om toch het aantal heropnamen in andere ziekenhuizen te schatten bij patiënten die zijn verhuisd, veronderstellen wij dat het gemiddelde aantal heropnamen per patiënt vergelijkbaar is tussen verhuisde en niet-verhuisde patiënten. Dit betekent dat het totale aantal heropnamen voor niet-verhuizers zou liggen rond de  $81.917/93.514 * 11.375 = 9.964$  heropnamen. In totaal konden wij 7.847 heropnamen herkennen: het aantal gemiste heropnamen bedraagt derhalve 2.117. Dit zijn de heropnamen van patiënten die én verhuisden én elders werden heropgenomen. Dit aantal geeft tevens een indicatie van de potentiële meerwaarde van het gebruik van de GBA tijdens de follow-up.

Een probleem bij bovenstaande berekening is echter het verschil in sterfte onder verhuizers en niet verhuizers. Dit hangt samen met de hoge sterfte kort na ontslag (zie figuur 4). Patiënten die kort na ontslag overlijden, hebben geen kans om te verhuizen. Het merendeel van deze vroege overlijders komt dan ook terecht in de groep niet-verhuizers. Dit houdt in dat het effectief te volgen aantal patiënten voor het optreden van heropnamen onder niet-verhuizers lager is dan de aanvangsgrootte doet vermoeden. Onderstaande tabel laat het fenomeen van de 'oversterfte' binnen de groep niet-verhuizers zien.

**Tabel 14.** Sterfte onder patiënten die wel en niet verhuisd zijn tijdens de follow-up.

Vitale status	Niet verhuizers (n=93.514)	Verhuizers (n=11.375)	Totaal (n=104.889)
Overleden	15.854 (17%)	957 (8,4%)	16.811 (16%)
In leven	77.660 (83%)	10.418 (92%)	88.078 (84%)

Om te corrigeren voor de vertekening die ontstaat door het verschil in sterfte hebben wij dezelfde berekening herhaald, maar nu met als noemer de patiënten die tot het einde van de follow-up in leven zijn gebleven in beide groepen (zie tabel 14). Het verwachte aantal heropnamen komt dan uit op  $81.917/77.660 * 10.418 = 10.989$ . Het

aantal gemiste heropnamen komt nu uit op 3.142 (10.989 minus 7.847). Helaas waren wij niet in staat om de heropnamen van patiënten die later overleden ook uit de teller te verwijderen. Hierdoor overschatten wij het aantal heropnamen in de niet-verhuizers groep. De beste schatting van het aantal gemiste heropnamen zal derhalve ergens liggen tussen de twee gemaakte schattingen  $(2.117 \text{ en } 3.142)/2 =$  ongeveer 2.500 heropnamen. Dit betekent dat na verhuizen ongeveer de helft van alle heropnamen bij dit algemene LMR-cohort weer plaatsvindt in het index-ziekenhuis. De andere helft wordt gemist tenzij veranderingen in de postcode kunnen worden opgepikt.

Het belang van verhuizen is mogelijk groter dan uit bovenstaande berekeningen naar voren komt. Zo vonden wij 1.167 heropnamen (via LMR-PAT-ID) waarbij de postcode was veranderd ten opzichte van de index-opname, maar geen bijbehorend verhuisbericht in de GBA. Hoogst waarschijnlijk betreft het hier patiënten die in een verpleeghuis/revalidatiekliniek zijn opgenomen, waarbij de ziekenhuisregistratie eerder de 'nieuwe' postcode gebruikt dan de GBA. Sowieso vraagt het punt van de synchroniciteit tussen LMR en GBA bij het veranderen van de postcode extra aandacht, wanneer de GBA tijdens de follow-up wordt gebruikt voor het up-to-date houden van de postcode.

Op grond van de verschillende definities is het ook mogelijk om een schatting te maken van het aantal heropnamen in hetzelfde ziekenhuis, maar waarbij LMR-PAT-ID (waarschijnlijk) niet goed is toegepast. Hiertoe hebben wij alle heropnamen in het index ziekenhuis van niet-verhuizers (op basis van hun unieke primaire sleutel) en alle heropnamen in het index-ziekenhuis van verhuizers, maar voor het tijdstip van verhuizen (ook op grond van hun initiële primaire sleutel) samengenomen. In totaal waren dit 70.730 heropnamen (eerste rij van tabel 13). Bij 63.970 (90%) van deze heropnamen stemde ook het LMR-PAT-ID overeen. Dit aantal valt alleszins mee daar bekend was dat 23 ziekenhuizen geen correct gebruik maken van het LMR-PAT-ID systeem (zie paragraaf 2.3 van hoofdstuk 2 voor uitleg).

## 5. DISCUSSIE EN CONCLUSIES

In deze pilotstudie zijn de mogelijkheden in Nederland onderzocht voor het achterhalen van heropnamen en/of overlijden in cohortonderzoek via registraties. De veronderstelling was dat groepen patiënten op nationaal niveau kunnen worden gevolgd door het combineren van gegevens uit drie landelijke registraties, te weten de Landelijke Medische Registratie, de Gemeentelijke Basisadministratie en de Statistiek van Doodsoorzaken. De mogelijkheden voor epidemiologisch onderzoek in Nederland voor het beantwoorden van (nationale) vragen over ziekte en zorg zouden hierdoor fors toenemen. Uiteraard diende het combineren en koppelen van gegevens uit de verschillende registraties te geschieden binnen de bestaande juridische en privacy regels van de respectievelijke houders van deze registraties. De pilotstudie was erop gericht een aantal stappen bij het herkennen van heropnamen en overlijden nader te onderzoeken. De gevolgde aanpak kan als voorbeeld dienen voor het toevoegen van gegevens uit andere registraties in Nederland (zie ook figuur 1 in de inleiding).

Onderzoek in de nabije toekomst zal zich richten op patiënten die opgenomen zijn geweest wegens een bepaalde aandoening, dan wel op patiënten die een bepaalde operatie hebben ondergaan. De epidemiologische analyses betreffen het patroon van sterfte (timing en doodsoorzaken) gedurende een bepaalde periode na ontslag, dan wel het patroon van heropnamen in combinatie met overlijden. In het pilotonderzoek is een ad hoc LMR-cohort gedefinieerd dat model staat voor toekomstige (LMR-)cohorten. De discussie geeft een samenvattend oordeel over de haalbaarheid en kwaliteit om heropnamen en overlijden te herkennen via landelijke registraties. Naast deze meer technische aspecten dienen bij toekomstig cohort-onderzoek ook diverse, meer inhoudelijke keuzes te worden gemaakt. Hoewel geen formeel onderdeel van deze pilot bespreken wij deze onderwerpafhankelijke keuzes hier kort (paragraaf 5.1). Het herkennen van heropnamen en overlijden, het feitelijke onderwerp van de pilot, wordt nadien besproken (herkennen overlijden in paragraaf 5.2; herkennen heropnamen in paragraaf 5.3). Voor beide onderdelen worden de relevante uitkomsten samengevat en onderliggende problemen beschreven en, waar mogelijk, gekwantificeerd. Ieder onderdeel eindigt met een samenvattend oordeel.

### 5.1 KEUZES BIJ OPZET COHORT-ONDERZOEK

Bij cohortonderzoek naar de prognose van specifieke patiëntengroepen dienen diverse inhoudelijke keuzes te worden gemaakt. Uiteraard speelt hierin de onderzoeksvraag zelf een belangrijke rol. In deze paragraaf beschrijven wij in algemene zin de verschillende keuzes. Deze keuzes liggen op de volgende gebieden:

- A. Formatie begincohort
  - A.1 specificatie index-aandoening
  - A.2 specificatie kalenderperiode
  - A.3 additionele in- en uitsluitcriteria

- B. Duur van de follow-up
- C. Gebeurtenissen van interesse

In het kort zullen wij bovengenoemde punten bespreken, waarbij de nadruk zal liggen op de formatie van het begincohort.

### **Ad A. Formatie begincohort**

In de pilot gaan wij uit van een (algemeen) LMR-cohort. Het is echter niet noodzakelijk dat het begincohort voortkomt uit de LMR. Ook van cohorten (patiëntengroepen) buiten de LMR kan het patroon van heropnamen en overlijden worden geanalyseerd, mits (een representatief deel van) deze patiënten kunnen worden teruggevonden in de GBA. Bij de verdere bespreking gaan wij wel uit van een LMR-cohort.

#### *Ad A.1 Specificatie index-aandoening*

Binnen de LMR kunnen specifieke patiëntengroepen worden bepaald via selectie op hoofdiagnose(n) of op operatiecode(s), dan wel een combinatie van beide. De hoofdiagnose is die diagnose (ziekte, aandoening) die achteraf (bij ontslag) verantwoordelijk wordt geacht voor de betreffende opname. Bij het coderen van hoofddiagnosen wordt gebruik gemaakt van het systeem van de International Classification of Diseases Clinical Modification (ICD-9-CM). Het gebruik van klinische opnamen bij het identificeren van patiëntengroepen heeft belangrijke consequenties. Het proefschrift 'Registers in Cardiovascular Epidemiology' van JB Reitsma, in het bijzonder hoofdstuk 1, geeft een overzicht van deze problematiek.<sup>1</sup>

Voor de identificatie van patiënten die een specifieke operatie hebben ondergaan, kan gebruik worden gemaakt van het feit dat iedere operatie/interventie tijdens een klinische opname wordt vastgelegd in de LMR. Een (potentieel) voordeel van het gebruik van operatiecodes is dat het leeuwendeel van de operaties plaatsvindt in het ziekenhuis, waardoor geen vertekening optreedt door selectieve uitval.

#### *Ad A.2 Specificatie kalenderperiode*

Naast de keuze van de index-diagnose(s) dient een besluit te worden genomen uit welke periode (kalenderjaar dan wel jaren) de patiënten uit het begincohort afkomstig zijn. Gezien de landelijke dekking van de LMR zal meestal één kalenderjaar volstaan om voldoende patiënten in te sluiten. De keuze voor een jaar dat verder in het verleden ligt, heeft als voordeel dat meer jaren voor follow-up (direct) beschikbaar zijn. Het vroegste jaar dat kan fungeren als index-jaar is 1995, het startjaar van de GBA.

#### *Ad A.3 Additionele in- en uitsluitcriteria*

Bij onderzoek naar de prognose van aandoeningen is het van belang dat patiënten zich in een vergelijkbaar stadium van de ziekte bevinden. Bij een willekeurige dwarsdoorsnede, zoals bij een kalenderjaar, zal de groep zowel bestaan uit patiënten in het begin van hun aandoening (eerste opname) als patiënten in een meer gevor-

---

<sup>1</sup> Reitsma JB. Registers in epidemiological research. In : Registers in cardiovascular epidemiology. Proefschrift, Amsterdam, 1999.

derd stadium (zoveelste heropname). Deze laatste groep van patiënten is waarschijnlijk verantwoordelijk voor de hoge sterfte kort na ontslag. Dit probleem kan (gedeeltelijk) worden ondervangen door alleen patiënten met een eerste opname in te sluiten via het screenen van voorafgaande LMR-jaren voor eerdere opnamen wegens dezelfde index-aandoening. Een alternatief is om deze informatie over eerdere heropnamen mee te nemen in de analyse. Ook kan informatie over andere prognostische factoren beschikbaar komen door te kijken naar eerdere opnamen, bijvoorbeeld voor het vaststellen van diabetes mellitus of niet. In technische zin verschilt het zoeken naar eerdere opnamen niet van het herkennen van vervolgoptnamen. Wel zijn de mogelijkheden voor het herkennen van eerdere opnamen voor 1995 beperkt: omdat de GBA pas in 1995 is gestart kan dit alleen binnen de LMR plaatsvinden en alleen voor opnamen vanaf 1991 (vanaf dit jaar wordt het 4-cijferige deel van de postcode in de LMR vastgelegd).

### **Ad B. Duur van de follow-up**

Een voldoende lange duur van de follow-up is van belang om een juist beeld te krijgen van de prognose van een ziektebeeld. De onderzoeksvraag en de index-aandoening bepalen in belangrijke mate wat een voldoende lange follow-up is. De kwaliteit van het herkennen van heropnamen zal afnemen met een langere follow-up. Een langere follow-up betekent namelijk meer verhuizingen die correct en eensluidend in LMR en GBA moeten worden verwerkt. Daarnaast neemt de stabiliteit van LMR-PAT-ID in de tijd af door introductie van nieuwe systemen van uitgifte van patiëntnummers, zoals bij fusies tussen ziekenhuizen. Voor overlijden (overlevingscurven) neemt de kwaliteit van de gegevens niet af met de duur van de follow-up.

### **Ad C. Gebeurtenissen van interesse**

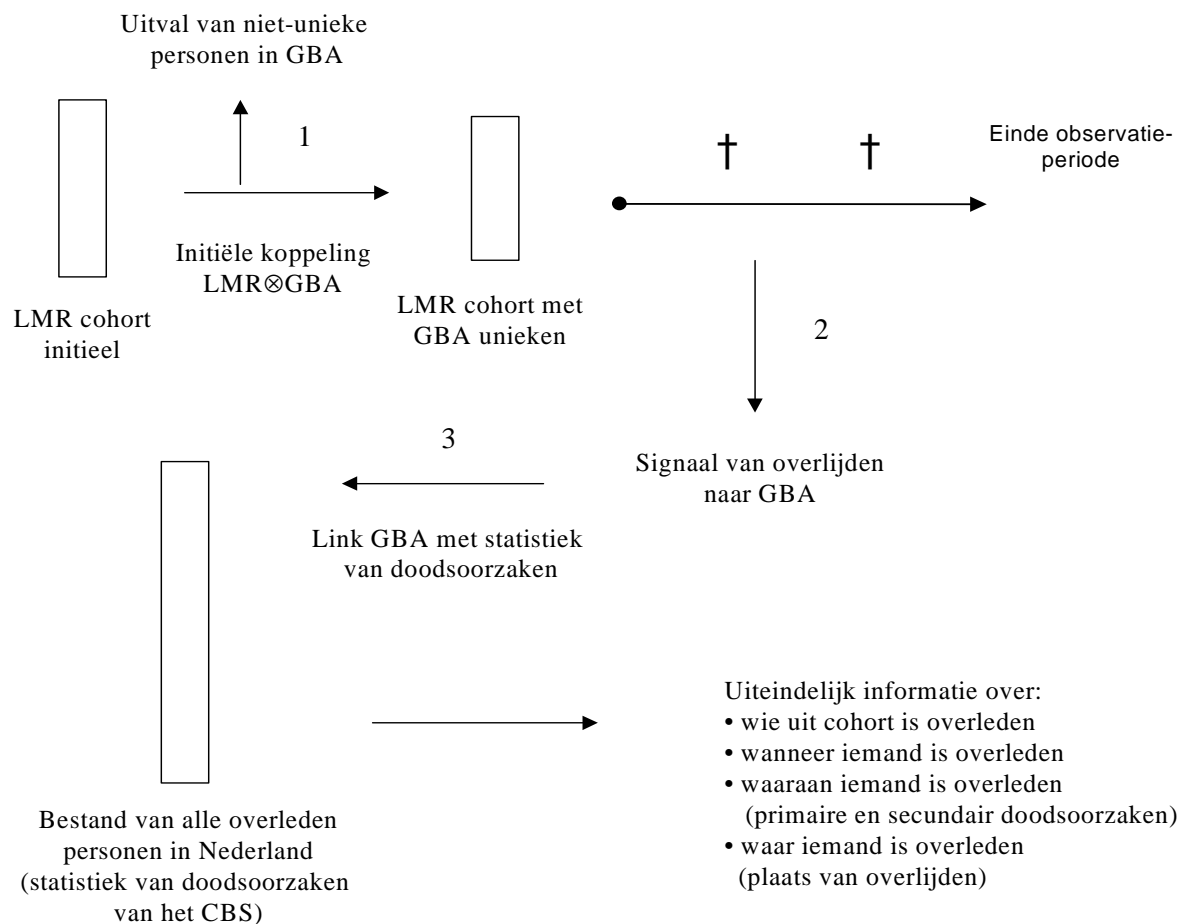
Bij het herkennen van heropnamen zijn diverse opties mogelijk. In de omvangrijkste variant wordt getracht alle heropnamen, ongeacht de hoofddiagnose dan wel operatiecode, te achterhalen. Dit betekent per follow-up jaar het doorzoeken van  $\pm 1,5$  miljoen ziekenhuisopnamen. Door vooraf te selecteren op bepaalde aandoeningen, bijvoorbeeld hart- en vaatziekten, neemt het aantal te doorzoeken opnamen sterk af. Dit bespaart niet alleen computertijd, maar maakt de koppeling ook overzichtelijker. Uiteraard worden eventuele heropnamen buiten de hart- en vaatziekten gemist.

Na deze bespreking van algemene keuzemomenten keren wij terug naar het eigenlijke onderwerp van de pilot: de technische mogelijkheden en problemen bij het herkennen van heropnamen en overlijden. Allereerst beschrijven wij het volgen van personen tot aan overlijden.

## **5.2 HERKENNEN VAN OVERLIJDEN**

Het herkennen van de sterfte dient zowel het overlijden binnen als buiten het ziekenhuis te omvatten. Figuur 5 geeft een schematisch overzicht van de verschillende

stappen van dit traject. Meer specifieke informatie over het herkennen van overlijden is terug te vinden in hoofdstuk 2 onder de paragraaf 2.1 'herkennen van overlijden'.



**Figuur 5.** De verschillende stappen en koppelingen bij het volgen van patiënten tot aan overlijden.

Op drie punten (zie ook nummering in grafiek) kunnen in dit traject problemen ontstaan. Deze zijn:

1. het niet, het niet-uniek, dan wel het foutief terug vinden van personen uit het initiële LMR-cohort in de GBA
2. het overlijden van personen uit het cohort zonder registratie in de GBA
3. het overlijden van een persoon uit het cohort met registratie in de GBA, maar zonder een corresponderend record in de statistiek van doodsoorzaken

Het gehele traject van het herkennen van overlijden is technisch uitgevoerd. Dit houdt in dat wij de sterfte van alle patiënten die uniek koppelden in de GBA hebben achterhaald en geanalyseerd (zie survival curve in figuur 4). In de pilotstudie heeft het empirisch onderzoek zich vervolgens beperkt tot probleem 1 dat zich afspeelt bij de initiële koppeling van LMR met GBA. Volgens CBS informatie (intern onderzoek)



komt probleem 2 vrijwel niet voor (rekening houdend met emigratie) is probleem 3 zeldzaam (beneden de 2%).

### *Initiële koppeling LMR\*GBA*

Op basis van drie gemeenschappelijke variabelen in LMR en GBA, te weten geslacht, geboortedatum en 4-cijferige deel postcode, waren wij in staat 84% van de LMR-patiënten uniek terug te vinden in de GBA. Bij ongeveer 11% van de cohortleden bleek deze combinatie van waarden *niet uniek* te zijn. Daarnaast was ongeveer 5% van de initiële cohortleden *niet terug te vinden* in de GBA. Deze 'niet terug te vinden' personen in de GBA betreffen mogelijk buitenlanders die in Nederlandse ziekenhuizen zijn opgenomen, personen die (bewust) niet, of niet juist zijn ingeschreven in de GBA en administratieve verschillen tussen LMR en GBA, dan wel fouten in LMR of GBA. Een mogelijk bron van problemen is het verschillend omgaan van LMR en GBA met 'tijdelijke' adressen, zoals bijvoorbeeld bij het opnemen van iemand vanuit een revalidatiekliniek of verpleeghuis. De anonieme basis van de koppeling maakt het uiteraard niet mogelijk binnen dit bestek hierover verdergaande uitspraken te doen.

Het niet-uniek terugvinden van een patiënt uit het cohort in de GBA is problematisch omdat dan meerdere personen kunnen worden gevolgd. Indien één van deze personen overlijdt is het onzeker of het handelt om de patiënt uit het cohort dan wel zijn of haar administratieve dubbelganger. Een directe oplossing is het uitsluiten van patiënten die niet uniek zijn terug te vinden in de GBA. Dit verlies gaat uiteraard ten koste van de power van het onderzoek. Belangrijker is echter of het volgen van uitsluitend unieke personen leidt tot vertekening (bias). Vertekening zal ontstaan wanneer unieke en niet-unieke personen van elkaar zouden verschillen in prognostische factoren met als gevolg differentiële sterfte. Voor een aantal potentiële prognostische factoren hebben wij naar hun invloed kunnen kijken. In het algemeen waren deze verschillen klein en lieten zij het volgende beeld zien. Unieke personen zijn gemiddeld ouder, wonen vaker in een postcodegebied met een kleiner aantal inwoners en zijn minder vaak allochtoon. Voor al deze variabelen geldt dat conditionele analyses mogelijk zijn, hetgeen bij voldoende aantallen correctie voor de betreffende vertekening mogelijk maakt.

Het probleem dat een LMR-patiënt uniek koppelt met een persoon in de GBA maar met de verkeerde, lijkt verwaarloosbaar. Bij minder dan 1% van alle patiënten die uniek koppelden vonden wij aanwijzingen dat deze koppeling incorrect was.

Ook na correcte identificatie in de GBA kunnen problemen optreden. Jaarlijks 'verdwijnt' ongeveer 0,2% van de personen uit het zicht van de GBA. Naast emigratie is het niet opgeven van een nieuw woonadres de belangrijkste reden voor dit verdwijnen.

## Conclusie over het herkennen van overlijden

Het herkennen van overlijden op landelijke schaal binnen cohort-onderzoek via bestaande registraties is technisch goed uitvoerbaar en vrijwel volledig. Voorwaarde hierbij is het (uniek) terugvinden van patiënten uit het cohort in de GBA, omdat langs deze weg zowel de sterfte binnen het ziekenhuis ( $\pm 40\%$  van de totale sterfte) als daarbuiten kan worden herkend. Door het geringe aantal gemeenschappelijke variabelen in LMR en GBA is ongeveer 85% van alle opgenomen Nederlanders uniek terug te vinden in de GBA. Het uniek dan wel niet-uniek zijn lijkt echter niet samen te hangen met een hogere dan wel lagere kans op sterfte. De onderzochte verschillen (leeftijd, grootte postcodegebied en etniciteit) waren klein, terwijl in de meeste gevallen correctie mogelijk is. Het overlijden wordt zelden gemist voor mensen die zijn ingeschreven in de GBA.

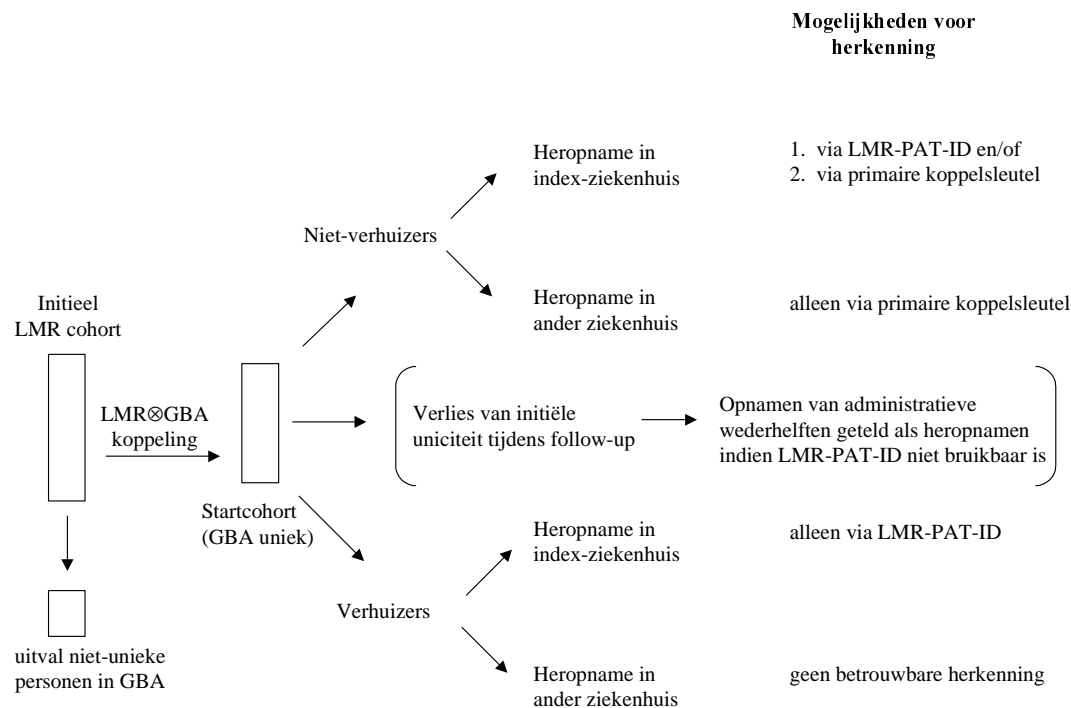
### 5.3 HERKENNEN VAN HEROPNAMEN

Het herkennen en analyseren van heropnamen binnen cohort-onderzoek is beduidend complexer dan het herkennen van overlijden. Dit heeft twee oorzaken.

Allereerst is bij het analyseren van heropnamen vaak ook het overlijden van personen uit het cohort belangrijk om het aantal personen 'at risk' correct te schatten. Dit houdt in dat ook een heropname-analyse start met een initiële koppeling met de GBA (uitzonderingen daargelaten), inclusief de daarbij horende problemen (zie paragraaf 5.2 '*herkennen van overlijden*').

Hier bovenop komt een tweede probleem: de sleutelrol van de variabele postcode bij het herkennen van heropnamen in de LMR (zie ook paragraaf 2.2 '*herkennen van heropnamen*' in hoofdstuk 2). Alleen bij heropnamen in hetzelfde ziekenhuis is herkenning mogelijk zonder tussenkomst van de postcode, mits het ziekenhuis correct gebruik maakt van het ziekenhuisspecifiek patiëntnummer. Dit houdt in dat heropnamen van personen die tijdens de follow-up verhuizen én in een ander ziekenhuis dan het index ziekenhuis worden heropgenomen, niet zijn te herkennen via LMR informatie alleen. Ook gaat het mis bij verhuisde patiënten die wel in hetzelfde ziekenhuis worden heropgenomen, maar waarbij het ziekenhuisspecifieke patiëntnummer niet correct wordt toegepast, zodat men alsnog op de inmiddels veranderde postcode is aangewezen. Een overzicht van de identificatiemogelijkheden van heropnamen zonder update van de postcode via de GBA tijdens de follow-up staat in figuur 6.

Door het toepassen van GBA informatie tijdens de follow-up is het mogelijk vanaf de (opgegeven) datum van verhuizing met de nieuwe postcode verder te zoeken. Daarnaast kan worden nagegaan of een unieke patiënt in de GBA ook uniek blijft tijdens de follow-up. Dit alles verhoogt de kwaliteit van de koppeling tussen LMR records onderling. Het continu gebruik maken van de GBA maakt het onderzoek wel ingewikkelder en arbeidsintensiever. Binnen de pilotstudie ontbrak de tijd en het geld om dit traject uit te voeren. Op basis van drie analyses is niettemin getracht een indruk te krijgen van de meerwaarde van de GBA tijdens de follow-up. Deze analyses zijn:



**Figuur 6.** Mogelijkheden en beperkingen bij het herkennen van heropnamen indien de GBA alleen initieel wordt gebruikt en niet tijdens de follow-up van patiënten. Het gebruik van andere partieel identificerende variabelen in de LMR (huisarts, betalende instantie, verzekeringswijze) is in deze figuur buiten beschouwing gelaten.

(1) een analyse van verhuisgegevens van alle Nederlanders en van een algemeen LMR-cohort; (2) een analyse naar het correct toepassen van LMR-PAT-ID binnen hetzelfde ziekenhuis; (3) een analyse naar de relatie verhuizen en heropnamen in een ander ziekenhuis dan van de index-aandoening.

### *Patroon van verhuizen in Nederland*

Wijzigingen in postcode (verhuizingen) zijn de motor achter het probleem van het missen van heropnamen. In het jaar 1996 verhuisde in totaal 7,5% van alle Nederlanders ten minste één keer waarbij het 4-cijferdeel van de postcode veranderde. Verreweg de hoogste verhuiskans (14,1%) was te vinden in de leeftijdsgroep 15-34 jaar. Voor alle andere leeftijden lag de kans op verhuizen rond de 5%. De jaarlijkse verhuiskans van LMR-patiënten week hier nauwelijks van af met 5,7% per jaar. Wel waren er aanzienlijke verschillen in verhuisgedrag per ziektebeeld (zie tabel 10).

### *Correct gebruik LMR-PAT-ID*

De meerwaarde van de GBA gedurende de follow-up komt ook naar voren bij het volgen van verhuizers die wel terugkeren naar het oorspronkelijke ziekenhuis, maar waar het LMR-PAT-ID niet goed wordt toegepast. Wij schatten het percentage heropnamen waarbij LMR-PAT-ID niet correct is toegepast op 10%. Deels heeft dit te maken met ziekenhuizen waarvan bekend is dat zij geen gebruik maken van LMR-PAT-ID (23 van de ongeveer 120 ziekenhuizen in Nederland in 1996).

### *Verhuizen en heropnamen in een ander ziekenhuis*

Binnen de pilotstudie is een schatting gemaakt van het aantal gemiste heropnamen in andere ziekenhuizen nadat patiënten zijn verhuisd. Dit levert tevens een schatting op van de (maximale) winst van het gebruik van de GBA gedurende de follow-up, mits deze optie foutloos zou werken. Om dit aantal te schatten zijn diverse aannames gemaakt, in het bijzonder dat de kans op heropname niet wezenlijk wordt beïnvloed door verhuizen (zie verder paragraaf 4.3.4). Ongeveer de helft van alle heropnamen na verhuizen worden gemist doordat deze plaatsvinden in een ander ziekenhuis. Uiteraard zal dit percentage afhangen van de onderliggende aandoening.

### **Conclusie over het herkennen van heropnamen**

Door de komst van de GBA zijn de mogelijkheden voor het herkennen van heropnamen binnen cohort-onderzoek sterk verbeterd. Het gebruik van GBA-gegevens lost namelijk twee problemen op:

1. het ontbreken van gegevens over sterfte buiten het ziekenhuis, zodat de juiste noemer (aantal personen at risk) onbekend bleef
2. het missen van heropnamen na verhuizen door de cruciale rol van de variabele postcode bij het herkennen van heropnamen

Een initiële koppeling met de GBA biedt een directe oplossing voor probleem 1 (zie conclusies aangaande het herkennen van overlijden). Daarnaast biedt deze initiële koppeling de mogelijkheid om te beginnen met een cohort GBA-unieke personen, hetgeen de kans verkleint om opnamen van administratieve wederhelften te tellen als heropnamen. De initiële koppeling met de GBA kent uiteraard dezelfde problemen als geschetst bij herkennen van overlijden: ongeveer 15% uitval door het niet uniek kunnen terugvinden van LMR-patiënten in de GBA.

Het tweede probleem kan worden gereduceerd door de GBA ook te gebruiken tijdens de follow-up om verhuizingen te herkennen en vervolgens de postcode aan te passen. Zonder update van de variabele postcode worden twee soorten heropnamen (in ieder geval) gemist in de follow-up van verhuizers: (1) heropnamen in een ander ziekenhuis dan het index-ziekenhuis; (2) heropnamen in hetzelfde ziekenhuis maar met een onjuist of niet beschikbaar LMR-PAT-ID. Al met al hangt het verlies dus af van: de jaarlijkse verhuiskans ( $\pm 7\%$  per jaar), de duur van de follow-up (hoe langer, hoe meer verhuizers), de 'trouwheid' aan het eerste ziekenhuis na verhuizen ( $\pm 50\%$  van de heropnamen is in hetzelfde ziekenhuis) en het niet kunnen terugvallen op LMR-PAT-ID (bij  $\pm 10\%$  van de heropnamen in hetzelfde ziekenhuis). Het gebruik van de GBA tijdens de follow-up is echter niet zonder problemen. Dit hangt vooral samen met de verschillende tijdstippen waarop in LMR en GBA de postcode wordt aangepast, vooral bij een al dan niet tijdelijke opname in een verpleeghuis. Daarnaast verdwijnen personen uit het zicht van de GBA, ongeveer  $0,2\%$  per jaar.

## 6. SAMENVATTING

De aandacht binnen epidemiologisch onderzoek verschuift steeds vaker van kortetermijn gevolgen van een ziekte naar gezondheidseffecten op de lange duur. Nationale registraties kunnen waardevolle longitudinale gegevens opleveren wanneer patiënten binnen en tussen registraties gevolgd zouden kunnen worden. Tot enkele jaren geleden waren deze mogelijkheden gering. In deze pilotstudie is binnen het CBS onderzocht of informatie uit de Gemeentelijke Basis Administratie (GBA) over mutaties door verhuizen en sterfte een rol kunnen spelen bij het in de tijd volgen van groepen patiënten.

Voor een cohort van ziekenhuisopnamen uit de Landelijke Medische Registratie (LMR, Prismant) in januari 1996 zijn de volgende vragen bestudeerd:

- ♦ hoe goed corresponderen gegevens uit de LMR met die uit de CBS-bevolkingsstatistieken?
- ♦ welke factoren beïnvloeden de koppeling tussen LMR en CBS-bevolkingsstatistieken?
- ♦ wat is het patroon van overlijden binnen dit cohort?

De koppeling geschiedde aan de hand van de combinatie geboortedatum, geslacht en het 4-cijferig postcodedeel. Van alle 124.598 opnamen uit het LMR-cohort kon 84,2% uniek worden toegekend aan een individu binnen de CBS-bevolkingsstatistieken. Bij 10,9% van de patiënten was deze koppeling te weinig specifiek; meerdere individuen uit de CBS-bevolkingsstatistieken koppelden met één patiënt uit de LMR. Bij 4,9% van de patiënten kon geen corresponderend bevolkingsstatistiek-record worden gevonden; deels zijn dit Nederland bezoekingende buitenlanders.

Voor het subcohort van 104.889 unieke patiënten kon met behulp van de CBS-bevolkingsgegevens eenvoudig het patroon van overlijden geanalyseerd worden na ontslag uit het ziekenhuis. De cumulatieve sterfte 2 jaar na ontslag varieerde van 44,4% voor patiënten met een maligniteit, 28,2% voor het acute hartinfarct, 56,6% voor hartfalen en 20,1% voor ziekten van de luchtwegen.

Door de komst van de GBA zijn de technische mogelijkheden voor het volgen van patiënten tussen en binnen registraties beduidend toegenomen. Uit deze pilot bleek dit met name uit het vrijwel volledig kunnen traceren van sterfte (ook buiten het ziekenhuis) en de grotere mogelijkheden voor het herkennen van heropnamen na verhuizen. Een voorwaarde is het uniek en correct terugvinden van een individu binnen de op de GBA gebaseerde bevolkingsstatistieken.

## BIJLAGE      FORMULES VOOR HET AANTAL ADMINISTRATIEVE MEERLINGEN

De volgende formules geven de verwachting en variantie van het aantal administratieve meerlingen in een steekproef zonder teruglegging. In een populatie van  $N$  personen is het aantal administratieve meerlingen en hun onderlinge opbouw bekend. De populatie is opgebouwd uit  $c_i$  groepjes van  $i$ -lingen. Elk  $c_i$  groepje bestaat uit  $i$  personen die zich  $i$ -ling mag noemen met  $i-1$  andere personen met dezelfde code, waarbij  $i$  loopt van 1 tot en met  $h$ .

*Voorbeeld:*

Code	1	2	3	4	5	6	7	8	9
Persoon	0	0	0	0	0	0	0	0	0
					0	0	0	0	0
								0	0

Hier geldt  $h=3$ ,  $c_1=4$ ,  $c_2=3$ ,  $c_3=2$  en  $N=1 \times c_1 + 2 \times c_2 + 3 \times c_3 = 16$

In het algemeen geldt dat  $1 \times c_1 + 2 \times c_2 + \dots + h \times c_h = N$

Trek nu aselect  $n$  personen uit de populatie  $N$  zonder teruglegging. Noem  $Z$  het aantal personen in de steekproef die minstens één codegenoot heeft in de steekproef. Kortom  $Z$  is het aantal meerlingen.

Definieer  $u_i = i \frac{\binom{N-i}{n-1}}{\binom{N}{n}}$ , voor  $i=1, \dots, h$

De verwachting van  $Z$  wordt nu gegeven door

$$E(Z) = n - \sum_{i=1}^h c_i u_i \tag{1}$$

In het speciale geval dat  $h=2$ ,  $c_1=0$  en  $2c_2=N$  gaat bovenstaande formule over in:

$$E(Z) = \frac{n(n-1)}{N-1} \tag{1a}$$

Bewijs:

$$E(Z) = n - \sum_1^h c_i u_i = n - (0 \times u_1 + \frac{N}{2} \times u_2) = n - \frac{N}{2} \frac{\binom{N-2}{n-1}}{\binom{N}{n}}$$

$$E(Z) = n - N \frac{\frac{(N-2)!}{(n-1)!(N-2-(n-1))!}}{\frac{N!}{n!(N-n)!}} = n - N \frac{(N-2)! n!(N-n)!}{(n-1)!(N-n-1)! N!} = n - N \frac{n(N-n)}{N(N-1)}$$

$$E(Z) = \frac{n(N-1)}{N-1} - \frac{n(N-n)}{N-1} = \frac{nN - n - nN + n^2}{N-1} = \frac{n^2 - n}{N-1} = \frac{n(n-1)}{N-1}$$

*Voorbeeld:*

N=15.000

waarvan  $c_1=12.000$  en  $c_2=1.500$

Stel we nemen een aselechte steekproef van 120 (n) personen. Dit is vergelijkbaar met een steekproef uit alleen de 'pool' van administratieve tweelingen ( $c_2$ ) van 24 (3.000/15.000\*120).

Op grond van formule (1a) met  $h=2$ ,  $c_1=0$  en  $c_2=1.500$ ,  $N=3.000$  en  $n=24$  levert een verwacht aantal administratieve meerlingen, in dit geval tweelingen, op van 0,184. Met andere woorden het aantal meerlingen neemt toe of af met het kwadraat van de steekproef. Indien een steekproef twee keer zo groot is neemt het aantal meerlingen een factor 4 toe. Oftewel het percentage meerlingen neemt lineair toe met de grootte van de steekproef.

*Alternatieve berekening*

Deze berekening gaat uit van de kans dat 2 willekeurige records dezelfde code hebben. Allereerst de kans dat 2 records op een specifieke code overeenstemmen. Deze kans is gelijk aan  $1/3.000 * 1/3.000 * 2 = 2/(3.000 * 3.000)$ . De kans op overeenstemmen op een willekeurige code is daarmee  $2/(3.000 * 3.000) * 1.500 = 1/3.000$

In een steekproef met grootte 24 is het aantal mogelijk paren gelijk aan  $\binom{24}{2}$

Kortom het verwachte aantal meerlingen is gelijk aan  $\binom{24}{2} \times \frac{1}{3000} \times 2 = \frac{24 \times 23}{3000} = 0,184$

Deze uitkomst is vergelijkbaar met de exacte benadering van hiervoor.