



# Working Papers Steunpunt Demografie



---

## De Nationale Databank Mortaliteit

### Aanmaak van een databank voor onderzoek van differentiële sterfte naar socio-economische status en leefvorm

P. DEBOOSERE & S. GADEYNE

STEUNPUNT DEMOGRAFIE  
VAKGROEP SOCIAAL ONDERZOEK (SOCO)  
**VRIJE UNIVERSITEIT BRUSSEL**

Working Paper 1999-7

Onderzoek uitgevoerd in het kader van de contracten AGORA 40 met DWTC en FWOAL 110 met FWO.  
De data werden toegeleverd door het NIS en vormen de Nationale Databank Mortaliteit.

---

**Steunpunt Demografie**, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium

Tel: 32-2-629.20.40

Fax: 32-2-629.24.20

E-mail: [esvbalck@vub.ac.be](mailto:esvbalck@vub.ac.be)

Website: <http://www.vub.ac.be/SOCO/>

**Vakgroep Bevolkingswetenschappen**, Universiteit Gent, Sint-Pietersnieuwstraat 49, B-9000 Gent, Belgium

Tel: 32-9-264.42.41

Fax: 32-9-264.42.94

E-mail: [John.Lievens@rug.ac.be](mailto:John.Lievens@rug.ac.be)

Website: <http://www.psw.rug.ac.be/dephome/bevowet>

## Inhoudstafel

1. Inleiding
2. Tellers en noemers
3. Vormt een Nationale Databank Mortaliteit geen inbreuk op de privacy?
4. De volkstelling van 1991 met het Rijksregister als basis
5. Wat is de kwaliteit van de koppeling tussen Volkstelling en Register?
6. De statistiek van de doodsoorzaken
7. Probabilistische koppeling met de statistiek van de doodsoorzaken
8. De koppelingsmethode
9. Gebruik van controlevariabelen
10. Finaal bestand
11. Overzicht van de covariaten
12. Besluit

## 1. Inleiding

De Nationale Databank Mortaliteit kwam tot stand in het verlengde van de werkzaamheden rond de resultaten van de volkstelling van 1991. Voor het eerst werd de volkstelling aangemaakt op basis van het Rijksregisternummer, een uniek identificatienummer voor elke inwoner van het rijk. De bedoeling was een aantal gegevens uit de klassieke volkstelling op die manier te betrekken uit het nationale bevolkingsregister. Dit was m.n. het geval voor de migratiegegevens. De succesvolle toepassing van de koppeling en exploitatie van volkstellings- en administratieve gegevens kende een logisch vervolg in de aanmaak van een Nationale Databank Mortaliteit. De mogelijkheid om volkstellingsgegevens te verbinden met het Rijksregister impliceert immers dat men op basis van het rijksregister de sterfte kan registreren van alle personen die in de volkstelling waren opgenomen.

Ondanks een traditiegetrouw uitstekend statistisch apparaat, scoort België internationaal eerder zwak op het vlak van mortaliteitsonderzoek. De reden hiervoor ligt niet enkel in de moeilijkheden van het afgelopen decennium inzake de verzameling van gegevens uit de overlijdenscertificaten. Ook het feit dat aan de sterftegegevens nauwelijks betrouwbare covariaten gekoppeld zijn speelt hier een belangrijke rol. Bovendien is het niet alleen belangrijk de covariaten van de overleden personen te kennen maar ook deze van de volledige bevolking om sterfterisico's correct te berekenen.

De volkstelling is de belangrijkste en enige exhaustieve bron van informatie over de socio-economische kenmerken van de bevolking in ons land. Deze informatie werd gekoppeld aan de registratie van overlijden in het Rijksregister voor de periode van 1 maart 1991 tot 1 maart 1996 en aan de informatie uit de overlijdenscertificaten tot eind 1994. De koppeling van deze drie bestaande databanken levert heel wat nieuwe informatie op m.b.t. differentiële sterfte en gezondheid van de bevolking in België en opent een brede waaier aan nieuwe studiemogelijkheden.

In het kader van het Agora-programma van de Federale Diensten voor Wetenschappelijke, Technische en Culturele Aangelegenheden werkt het Steunpunt voor Demografie aan een pilootproject. Dit gebeurt in opdracht van en in samenwerking met het Nationaal Instituut voor de Statistiek. Het Agora project « Nationale Databank Mortaliteit » heeft tot doel de bestaande administratieve data te exploreren op hun potentieel voor de analyse van de mortaliteit naar socio-economische kenmerken. Deze verkennende analyse zou optimaal moeten uitmonden in het opzetten van een permanente databank.

Het onderstaand artikel heeft als doel deze nieuwe databank voor te stellen. We geven een overzicht van de 3 bronnen waaruit de databank bestaat en bespreken de gevolgde werkwijze bij hun onderlinge koppeling. Tevens worden de mogelijkheden en de beperkingen van de nieuwe databank gedocumenteerd.

## 2. Tellers en noemers

Centraal in de ontwikkeling van de databank mortaliteit staat het probleem van de “tellers en de noemers”. Het sterfteregister geeft ons een aantal inlichtingen over de overleden personen zoals leeftijd en beroep. Aan de hand van die gegevens kan een tabel worden aangemaakt van bijvoorbeeld de mortaliteit per beroepscategorie. Maar deze informatie vertelt eigenlijk niets over sterfterisico's als niet tegelijk de verdeling gekend is van de beroepscategorieën over de gehele bevolking. Het is precies deze lacune die de Nationale Databank Mortaliteit opvult. Waar onze belangstelling naar uitgaat is de verhouding tussen het aantal overlijdens en de totale risicopopulatie voor een gegeven eigenschap.

De bevolking zoals geregistreerd in de volkstelling van maart 1991 vormt de totale risicopopulatie en levert de noemers. De sterfte uit het Rijksregister of uit de sterfteregisters geeft de gebeurtenissen en dus de tellers. Alleen wanneer beide samenkomen kunnen conditionele kansen worden berekend. Is binnen een gegeven periode de kans op overlijden van een persoon met een lagere opleiding groter deze van een hoger opgeleide? Is de overlijdenskans van een man tussen 35 en 64 jaar al dan niet afhankelijk van zijn tewerkstelling?

Kansberekening kan natuurlijk ook op basis van geaggregeerde gegevens. Wanneer het aantal overlijdens gekend is voor een bepaalde periode voor een gewest dan kan dat worden afgezet tegen de risicopopulatie in dat gewest. Deze methode heeft echter twee belangrijke nadelen. Wanneer de sterfteregistratie niet op individuele basis verbonden is met de risicopopulatie dan komt de relatie tussen tellers en noemers heel snel op losse schroeven te staan. Bij de registratie van de sterfte over een aantal jaren is er na een bepaalde tijd helemaal geen zekerheid meer dat de risicopopulatie nog overeenstemt met de aanvangspopulatie. Verhuisbewegingen kunnen de samenstelling van de bevolking, zeker in een kleinere regio, grondig wijzigen in de loop van een beperkt aantal jaren. En die wijzigingen kunnen zeer belangrijk zijn wanneer het gaat om de berekening van sterfterisico's bij kleine subpopulaties. Alleen een databank die risicopopulatie en sterfte individueel met elkaar verbindt kan ook rekening houden met wijzigingen in de risicopopulatie door migratie.

Een tweede nadeel van het gebruik van geaggregeerde data is het volledig gebrek aan controle op de onderlinge relatie van de verschillende variabelen en hun invloed op het sterfterisico. Het aanmaken van een bestand met individuele gegevens geeft het grote voordeel dat individuele sterftetekansen kunnen worden berekend voor een bepaalde variabele rekening houdend met andere variabelen (multivariate analyse). Differentiële sterfte tussen gewesten kan worden gecontroleerd op verschillen in beroepsstructuur, inkomen, opleiding enz. of m.a.w. er kan getoetst worden of regionale differentiatie in sterfte al dan niet samenhangt met verschillen in socio-economische samenstelling van de bevolking. De aanmaak van een Nationale Databank Mortaliteit licht het deksel op van de black box waarbinnen het onderzoek met geaggregeerde data gevangen zat.

### 3. Vormt een Nationale Databank Mortaliteit geen inbreuk op de privacy?

Bij de ontwikkeling en het gebruik van databanken komt onvermijdelijk het probleem van de bescherming van de privacy aan de orde. In het licht van de opdracht van het Nationaal Instituut voor de Statistiek ligt het voor de hand dat deze instelling probeert de databanken die zij beheert beter te exploiteren en te rentabiliseren. Het aanmaken van een databank zoals een volkstelling of een huishoudensbudgetenquête is een zeer dure aangelegenheid. Het is dus aangewezen de gegevens die men op die manier heeft verzameld optimaal te laten renderen. Het onderling verbinden van databanken is daarvoor de aangewezen methode, en laat toe om nieuwe dimensies toe te voegen aan een bestaande databank tegen een relatief lage kostprijs. Het verbinden van de volkstelling met de migratiegegevens uit het Rijksregister laat bijvoorbeeld toe migratiebewegingen te analyseren naar socio-economische kenmerken.

Een ander voordeel is dat men bestaande databanken een nieuw leven inblaast of up to date houdt door ze te verbinden met actuele data. Zo zijn de achtergrondvariabelen van de volkstelling van 1991 nog steeds nuttig om migratiebewegingen in 1996 te analyseren of, in casu, onderzoek te doen naar differentiële sterfte.

Het ontstaan van geïnformatiseerde administratieve databanken laat ook toe heel wat statistische gegevens aan te maken. Zo is de statistiek van de loop van de bevolking sinds 1988 gebaseerd op de gegevens van het Rijksregister. Het onderling koppelen van administratieve databanken zoals nu reeds gebeurt door de Kruispuntendatabank vormt ook hier het logisch vervolg voor het rationeel aanmaken van statistieken.

Het grootste bezwaar tegen deze evolutie is ongetwijfeld de vrees dat dergelijke meta-databanken een ernstige inbreuk kunnen vormen op het privé-leven. Het is vanuit die angst dat ook de wetgeving rond de bescherming van de privacy de jongste jaren terecht sterk is geëvolueerd. Vandaar het belang om, bij het aanmaken van databanken voor wetenschappelijk onderzoek, op dit punt met de grootste omzichtigheid tewerk te gaan. In het kader van deze opdracht voor het N.I.S. zijn de medewerkers van het Steunpunt voor Demografie gebonden door het statistisch geheim. Dit belet niet dat alles in het werk werd gesteld om elke individuele herkenning uit te sluiten. Alle gegevens zijn uiteraard *geanonimiseerd*: naam- en adresgegevens werden aan de bron op het N.I.S. weggelaten. Verder werd op geen enkel moment met het Rijksregisternummer gewerkt omdat dit nummer toegang kan geven tot andere administratieve informatie. De omzettingstabellen tussen Rijksregister en volkstelling werden gescheiden gehouden van de eigenlijke databank en de omzetting zelf gebeurde op het N.I.S.. Bij verspreiding en publicatie van de gegevens geldt bovendien dat het aggregatieniveau dusdanig moet worden gekozen dat ook individuele herkenning op basis van de kruising van variabelen uitgesloten is. Hiermee is het principe van de *niet-identificeerbaarheid* toegepast.

#### 4. De volkstelling van 1991 met het Rijksregister als basis

Op de kwaliteit van de gegevens van de volkstelling zelf wordt hier niet ingaan. Daarvoor verwijzen we naar andere publicaties die aan dit thema uitgebreid aandacht hebben besteed (zie o.m. Stoop en Surkyn). Wel is het nodig om hier in te gaan op de invloed van het gebruik van het Rijksregister voor de volkstelling van 1991. Op basis van de gegevens van het Rijksregister werden de telformulieren aangemaakt en opgestuurd naar de betrokkenen. Het Rijksregister van de natuurlijke personen is een gecentraliseerd en geïnformatiseerd bevolkingsregister. De wet op het Rijksregister van 8 augustus 1983 legt de gemeenten op om voor elke bewoner van de gemeente 9 basisgegevens aan het register over te maken. Het Rijksregister bevat aldus voor iedere inwoner volgende gegevens :

1. de naam en voornamen
2. de geboorteplaats en de geboortedatum
3. het geslacht
4. de nationaliteit
5. de hoofdverblijfplaats
6. de plaats en datum van overlijden
7. het beroep
8. de burgerlijke staat
9. de samenstelling van het gezin.

Het Rijksregister biedt het voordeel op elk moment de samenstelling van de bevolking te kunnen geven voor elk administratief niveau. Door het Rijksregister te gebruiken was het mogelijk de telformulieren per post op te sturen en ook reeds de samenstelling van het huishouden voor te drukken op het zogenaamde gezinstelformulier. Het Rijksregister van 1 december 1990 vormde hiervoor de basis. Verbeteringen dienden te worden aangebracht indien de situatie tegen 1 maart 1991 gewijzigd was (overlijden, geboorte, migratie) of indien de registratie foutief bleek te zijn.

Het gebruik van het Rijksregister bij de volkstelling van 1991 heeft een aantal belangrijke implicaties. Er zijn onbetwistbare voordelen. Het vertrekken van de bestaande administratieve databank laat toe een aantal gegevens zeer accuraat in te brengen (geboortedatum, datum van inschrijving, enz.). Het geeft ook de mogelijkheid controles uit te voeren door terugkoppeling. Zo kunnen bijvoorbeeld onwaarschijnlijke antwoorden worden gecorrigeerd. Tenslotte is het mogelijk een aantal gegevens niet langer te bevragen maar uit de administratieve databank te putten (huwelijksdatum, duur van verblijf, verhuisbewegingen). We moeten echter ook de beperkingen inschatten. De basis is een legale, officieel geregistreerde situatie waar de praktijk soms van afwijkt. Bij controle van de Rijksregistergegevens naar aanleiding van de volkstelling in de gemeente Etterbeek bleek er een afwijking te bestaan van maar liefst 11% t.o.v. de officieel ingeschreven bewoners. De bevolking van Etterbeek zou in 1991 aldus circa 4000 personen meer tellen dan de de jure bevolking van 38.000 (Census Belgica 2001, 1997, p.8). Alleen Etterbeek voerde een dergelijke operatie uit naar aanleiding van de volkstelling. Voor de rest van het land was het vertrouwen in het Rijksregister vrij groot en waren de aanpassingen doorgaans minimaal.

De telformulieren werden via de post opgestuurd maar dienden door tellers te worden opgehaald. De instructies van het N.I.S. voor deze ophaling zijn duidelijk en gedetailleerd

uitgeschreven in een handleiding voor de tellers. “De teller verifieert of de telformulieren ondertekend zijn en weigert de niet- of duidelijk onvoldoende ingevulde telformulieren.” (N.I.S., 1991a, p.14). De meeste telplichtigen en de meeste tellers hebben dit correct gedaan. Desondanks is een deel van de formulieren onvoldoende gecontroleerd en beperken de ingevulde vragen zich tot de voorgedrukte gegevens uit het Rijksregister. Bewijs hiervan zijn de formulieren van personen die bij de volkstelling al niet meer tot de populatie behoorden. Zo stelden we bij de aanmaak van deze databank vast dat niet minder dan 97 personen in de volkstelling al overleden waren vóór de datum van de telling en 471 waren vóór die datum reeds geëmigreerd.

Op die manier werden formulieren die in werkelijkheid niet door de bevrageden werden ingevuld toch in de volkstelling opgenomen. Bij een studie van de non-respons onder de bevolking van 18 tot 60 jaar kwamen we tot de vaststelling dat 3,4% van de individuele tellingformulieren niet werden ingevuld. Voor 79% van deze gevallen ontbreken ook de gegevens van de woningen (Stoop en Surkyn, 1997). We kunnen er dus van uitgaan dat voor 2,7% van de bevolking tussen 18 en 60 jaar de gegevens louter uit het Rijksregister werden geput. Deze niet ingevulde tellingformulieren zijn vanzelfsprekend niet toevallig verdeeld over de populatie. Stoop en Surkyn komen tot de conclusie dat geheel ontbrekende gegevens vooral voorkomen bij alleenstaanden, jongeren en vreemdelingen. Personen die hun tellingformulieren niet invullen zijn vooral niet toevallig verdeeld wat het sterfterisico betreft. Ontbrekende data (“missings”) of volledig ontbrekende records (“recordwise missings”) zullen onvermijdelijk hoger scoren qua sterfterisico dan de rest van de populatie. Het spreekt vanzelf dat mensen die geconfronteerd worden met een zware ziekte of die slachtoffer waren van een ernstig ongeval, en dus een groter sterfterisico vertonen, tegelijk ook veel meer kans maken om hun tellingformulier niet in te vullen. Daarenboven zullen de kenmerken van die groep sterk van andere factoren afhangen. Terwijl een alleenstaande in een dergelijke situatie zelden een formulier zal invullen, kunnen huisgenoten inspringen voor een gezinslid in moeilijkheden. Er is ook nog een andere relatie tussen missings en sterftcijfers. Marginale of gemarginaliseerde personen zullen voor beide een groter risico vertonen. M.a.w. handicaps op gezondheidsvlak of op maatschappelijk vlak – en vaak gaat het hier om een accumulatie van handicaps – zijn de gemeenschappelijke grondslag voor een hoger sterfterisico en een grotere kans op het niet of het slecht invullen van volkstellingformulieren. Bij de interpretatie van de analyse van de nationale databank mortaliteit is het zeer belangrijk met dit gegeven rekening te houden.

Een probleem dat hierbij aansluit betreft de codering van de “missings” door het Nationaal Instituut voor de Statistiek. Door het feit dat voor bepaalde variabelen geen specifieke codering werd gebruikt voor de “missings” worden deze soms gelijkgesteld met de code “nul”. Dit is bijvoorbeeld het geval met de informatie over de pariteit. Bij vrouwen waar de gegevens ontbreken komt hier de code “nul” voor, waaruit men zou kunnen afleiden dat zij geen kinderen hebben gekregen. In werkelijkheid gaat het om “recordwise missings” van vrouwen die misschien niet, maar misschien ook wel, kinderen hebben. Maar ook de hercodering van bepaalde variabelen kan de kwaliteit hier geweld aandoen. Ontbrekende gegevens worden door het N.I.S. soms aangevuld door informatie te putten uit andere wel ingevulde gegevens. Deze hercodering is vaak correct en verhoogt het aantal valabele antwoorden, maar in sommige gevallen ontstaat hier ook een ernstige ruis op de gegevens. Het is essentieel in de toekomst een onderscheid te

kunnen maken tussen gegevens die door de respondent werden ingevuld en gegevens die achteraf door hercodering werden aangevuld of gecorrigeerd.

We kunnen besluiten dat het gebruik van het Rijksregister als basis voor de telling bij deze eerste operatie wellicht geen grotere nauwkeurigheid heeft opgeleverd dan de volkstelling van 1981 en dat we moeten waakzaam zijn voor een vals gevoel van precisie dat door het gebruik van administratieve data wordt opgewekt. Toch blijft de volkstelling met een brede waaier aan socio-economische gegevens over de totale bevolking een unieke, buitengewoon rijke bron van informatie over de Belgische bevolking.

## 5. Wat is de kwaliteit van de koppeling tussen Volkstelling en Register?

In het kader van de volkstellingsmonografieën kreeg het Steunpunt voor Demografie de opdracht de koppeling tussen de volkstelling en het Rijksregister aan te maken, te valideren en te exploiteren, vooral met het oog op de analyse van de interne migraties. Voor deze analyse was het van belang niet alleen de migratiegegevens zelf te koppelen, maar ook de eventuele sterftedatum. Wat de basis vormde voor migratieonderzoek opende meteen het perspectief naar nieuwe onderzoeksmogelijkheden inzake mortaliteit. Voor de verwerking door het Steunpunt voor Demografie werden de bestanden van het Rijksregister vooraf geanonimiseerd. Zowel naamgegevens als adresgegevens werden verwijderd. Ook werd het Rijksregisternummer vervangen door het volkstellingsnummer op basis van het oorspronkelijke hercoderingsprogramma van het N.I.S.. Voor personen waar geen volkstellingsnummer voor bestond werd een willekeurig sequentiële nummer aangemaakt.

Er was dus een bestand afkomstig uit het Rijksregister met de gemeente van verblijf, het geslacht, de geboortedatum en in voorkomend geval de datum van emigratie of van overlijden. Voor elk van de vijf jaren die op de volkstelling volgen bestaat een dergelijk bestand. Aan de hand van het volkstellingsnummer was het mogelijk deze gegevens te koppelen aan de gegevens uit de volkstelling. De bevolking op datum van de volkstelling van 1 maart 1991 vormt dus het uitgangspunt. Dit bestand werd gekoppeld aan het bestand uit het Rijksregister van 1996 waarbij de toestand werd nagegaan op 1 maart 1996. Voor de grootste groep was het mogelijk de informatie uit het Rijksregister te koppelen aan de volkstelling. Een deel van deze mensen overleed of emigreerde in de eerste twee maanden van 1996. Voor hen kon men een preciese datum van emigratie of overlijden uit het Rijksregister putten. Maar er was ook een belangrijke groep personen die tot de volkstelling behoorden en waarvoor geen informatie meer beschikbaar was in het register. Voor deze groep werden, teruggaand in de tijd, de gegevens opgezocht in de respectieve registerbestanden van de tussenliggende jaren. Aldus was het mogelijk ook voor iedereen die emigreerde of overleed in de tussenperiode een preciese datum te registreren. Op die manier ontstaat een bestand waarbij voor elk van de relevante covariaten van de volkstelling zowel de sterfte als de risicopopulatie gekend zijn.

Het bestand van de volkstelling bevat 9.978.681 personen. Voor 9.329.881 personen was het mogelijk een koppeling te maken met het register van 1996. Voor diegenen die in de



loop van de eerste twee maanden van 1996 overleden of migreerden hebben we ook de desbetreffende datum.

Van de overblijvende 648.800 personen vonden we van 27 personen geen koppelingsgegevens meer. Voor de anderen werden de gegevens achtereenvolgens uit elk van de voorgaande registerjaren geput.

Rijksregister	Aantal gerecupereerde gegevens
1996	9.329.881
1995	131.029
1994	131.079
1993	133.749
1992	134.307
1991	118.609
Geen link	27

*Tabel 1: Overzicht van de records uit de volkstelling van 1991 en de koppeling met de Rijksregistergegevens. De eerste koppeling gebeurde op basis van het Rijksregister in 1996. De voorgaande jaren geven telkens het totaal aantal personen dat uit het Register verdween door overlijden of migratie.*

Het kleiner aantal records dat we uit het register van 1991 recupereren vloeit voort uit het feit dat we voor 1991 slechts 10 maanden overspannen vanaf de datum van de telling (1 maart 1991). Van al deze personen kennen we nu de exacte migratiedatum of overlijdensdatum. Een aantal personen die het land hebben verlaten hebben hun bestemming en datum van vertrek niet opgegeven. Ze werden door de gemeente ambtshalve geschrapt op het moment dat vastgesteld werd dat ze niet langer in de gemeente verbleven. Voor deze personen werd de datum van schrapping gelijkgesteld met de datum van migratie.

Controle of de terugkoppeling tussen Rijksregistergegevens en volkstelling wel correct is verlopen, gebeurde door een vergelijking van variabelen die in de twee bestanden voorkomen en die niet gebruikt werden om de koppeling te realiseren: geslacht en geboortedatum. Voor beide variabelen werd geen enkele afwijking vastgesteld waarbij kan worden gesteld dat de terugkoppeling voor 100% correct lijkt. Op de totale populatie waren er slechts 7 personen die in de volkstelling waren opgenomen en die we niet konden traceren in het Rijksregister.

Het is interessant de gegevens van de volkstelling en van het Rijksregister onderling met elkaar te vergelijken. Dit levert ons heel wat informatie op over de mogelijkheden om in de toekomst een klassieke census te vervangen door de loutere koppeling en telling van administratieve data.

Gezien de volkstelling werd uitgevoerd op basis van het Rijksregister en het register achteraf werd gecorrigeerd door de volkstelling is een goede overeenkomst vanzelfsprekend, althans voor de gegevens waarover we beschikten. Vermoedelijk is er een kleine onderschatting in het totaal bevolkingscijfer want noch het register, noch de

telling bevatten de gegevens over personen die niet officieel werden ingeschreven. Externe en interne validatie laten ook toe aan te nemen dat de gegevens m.b.t. geslacht, nationaliteit, burgerlijke stand en geboortedatum vrij correct zijn en in elk geval in beide bestanden vrijwel volkomen overeenstemmen. Wat de woonplaats betreft weten we dat we hier met officiële registraties te maken hebben. Om allerlei redenen van fiscaliteit of sociale wetgeving kan die soms afwijken van de realiteit. Maar we kunnen, minstens vanuit een administratief standpunt, een goed beeld krijgen van de migratiebewegingen. Onderzoek op de overige gegevens uit het register leert verder dat deze wel een aanzienlijke foutenmarge vertonen. De samenstelling van het gezin en vooral het beroep zijn in hoge mate onbetrouwbaar. Controles op de samenstelling van het gezin wezen uit dat deze informatie soms incompatibel was met andere gegevens zoals de verwantschap met de referentiepersoon en dat het Rijksregister soms pas jaren na een verandering van situatie een aanpassing doorvoert. Meer nog, een belangrijk percentage van de data in het register worden gewoon niet aangepast. Volgens de procedures die gelden voor het inbrengen van gegevens in het register zou dit normaal niet mogen voorkomen. Wellicht wordt deze procedure niet steeds nauwkeurig opgevolgd door alle gemeenten. Bovendien ontbreken nationale correctiemechanismen. Wat het beroep betreft is algemeen bekend dat dit hier hoogstens correct wordt genoteerd bij een eerste inschrijving in de gemeente. Van latere veranderingen in beroepssituatie wordt vrijwel nooit akte genomen.

## 6. De statistiek van de doodsoorzaken

De statistiek van de doodsoorzaken valt in ons land uiteen in twee bestanden. Een bestand bestaat uit de overlijdens van kinderen jonger dan één jaar. Het andere bestand omvat alle overige overlijdens. De laatste jaren is de beschikbaarheid van de statistiek van de doodsoorzaken met ettelijke jaren vertraagd. Dit is een rechtstreeks gevolg van een enigszins chaotische situatie die is ontstaan bij het overbrengen van de bevoegdheid voor deze materie van de Federale overheid naar de Gemeenschappen. Vooral de situatie in het Brussels Gewest zorgde daarbij voor problemen. Momenteel is door een goede samenwerking tussen de nieuwe diensten voor het Brussels Gewest en de Vlaamse Gemeenschap het einde van deze situatie in zicht, op voorwaarde dat de overheden in de diverse Gemeenschappen en Gewesten deze inspanningen en samenwerking daadwerkelijk verder steunen en uitbouwen.

Tot in 1990 bestonden twee bestanden in België met de gegevens van de formulieren voor aangifte van overlijden. Er was het bestand van het Ministerie van Volksgezondheid en Leefmilieu en het bestand van het Nationaal Instituut voor de Statistiek. Het Ministerie van Volksgezondheid en Leefmilieu centraliseerde tot in 1990 de formulieren die werden gecodeerd door de Franse en de Vlaamse Gemeenschap op basis van de ICD-9 (World Health Organization, 1977) Vanaf 1991 wordt het bestand op nationaal vlak enkel nog aangemaakt bij het NIS. De formulieren komen terecht zowel bij de Gemeenschappen als bij het N.I.S.. Beide instellingen brengen deze gegevens in, maar coördineren daarna wel hun inspanningen.

De overlijdenscertificaten bevatten vier stroken (A, B, C en D). Voor de preciese inhoud van deze stroken verwijzen we naar appendix 1. Welke weg leggen deze certificaten nu af alvorens te worden verwerkt in een centraal bestand dat alle Belgische overlijdensgegevens bevat? Doorgaans doorloopt een certificaat minimaal een vijftal stappen alvorens in een definitief computerbestand te worden opgenomen. Bij elk van die stappen kunnen fouten in het bestand sluipen.

#### *1. De geneesheer.*

Het overlijden van een persoon in België wordt vastgesteld door een geneesheer die een certificaat van overlijden dient op te stellen. Doorgaans is het de huisarts of de behandelende geneesheer die het certificaat maakt. Bij personen die in bijzondere omstandigheden overlijden (verdrinking, sommige gevallen van zelfmoord,...) wordt een arts opgeroepen om de dood vast te stellen. Artsen die zich als vrijwilliger hebben opgegeven, dienen evenwel geen enkele bijzondere kwalificatie of opleiding te hebben. Stroken A, B en C dienen te worden ingevuld door de geneesheer die het overlijden vaststelt. Strook C bevat de inlichtingen betreffende het overlijden en, meer bepaald, de doodsoorzaken. Deze strook wordt door de arts dichtgekleefd en zal alleen worden geopend door de bevoegde diensten wanneer de persoonsgegevens van strook A verwijderd zijn. Alleen strook A bevat naam en adres van de overledene. Deze strook is te bewaren door het gemeentebestuur en wordt niet aan andere instanties doorgegeven.

#### *2. De gemeente*

Het gemeentebestuur zorgt ervoor dat deze gegevens ook in het Rijksregister worden ingebracht. Alle overige besturen krijgen dus enkel de geanonimiseerde formuliergegevens. De gemeenteambtenaar moet wel alle gegevens controleren van de stroken A en B en moet de strook D invullen. Hiervoor wordt gebruik gemaakt van het bevolkingsregister van de gemeente en van het Rijksregister. De gemeente geeft verder alleen de stroken B,C en D door aan de Provinciale Gezondheidsinspectie; strook A blijft in de gemeente. Een bijkomende stap is vereist indien het overlijden plaats vindt in een andere gemeente dan de gemeente van verblijf. In dit geval dient de gemeente van overlijden het certificaat over te maken aan de gemeente van verblijf die de verdere administratieve stappen verwerkt.

#### *3. De Provinciale Gezondheidsinspectie*

Slechts in dit stadium wordt strook C geopend. Strook A met de identificatiegegevens is hier immers reeds verwijderd. De Provinciale Gezondheidsinspectie controleert strook C op ontbrekende gegevens.

#### *4. De Gemeenschappen*

De drie stroken worden vervolgens aan Vlaamse kant overgemaakt aan het Ministerie van de Vlaamse Gemeenschap waar een centraal coderingsteam instaat voor de codering. De strook C wordt bewaard bij de Vlaamse Gemeenschap. Voor Brussel komen deze gegevens sinds 1998 terecht bij het Centrum voor Gezondheidsobservatie dat de bewerking overlaat aan de coderingsploeg van de Vlaamse Gemeenschap. Het Brusselse Centrum voor Gezondheidsobservatie geeft vervolgens de respectieve overlijdensgegevens door aan de Vlaamse – en aan de Franse Gemeenschap. Voor de Franse Gemeenschap worden de gegevens voor het Waals Gewest afzonderlijk ingezameld.

#### *5. Het Nationaal Instituut voor de Statistiek*

De stroken B en D worden overgemaakt aan het N.I.S.. Voor de jaren 1991 tot 1994, welke werden gebruikt voor de aanmaak van de huidige databank, werden de gegevens

blijkbaar door het N.I.S. zelf op computer ingebracht. Op het N.I.S. worden de gegevens dan nogmaals gecontroleerd met een aantal eigen controleprogramma's. Tot 1998 werd deze laatste procedure eigenlijk parallel doorgevoerd met de verwerking van de gegevens door de Gemeenschappen. Er bestaan dus verschillende databanken naast elkaar die onderling ook nog eens kunnen verschillen. De laatste jaren werden stappen ondernomen om een einde te maken aan deze toestand en zou het N.I.S. in de toekomst de geïnformatiseerde bestanden van de Gemeenschappen overnemen. Een betere samenwerking zou de kwaliteit van de gegevens zeker ten goede komen. De verwarring van de afgelopen jaren heeft wat de nationale data betreft daar zeker niet toe bijgedragen. Vast staat dat een aantal hercoderingsregels van de WGO voor de onmiddellijke en oorspronkelijke sterfteoorzaak niet correct werden toegepast. Vergelijkend onderzoek naar de sterfteoorzaken tussen de Gewesten bracht verschillen aan het licht die eerder het gevolg zijn van een verschillende registratie- en/of codering dan wel van een verschil in werkelijke sterfteoorzaak. Bij de verdere verfijning van de databank zal op deze problemen zeker verder moeten worden ingegaan. Voorlopig hebben we onze analyse vooral toegespitst op de foutieve registratie van de administratieve variabelen in het sterftecertificaat.

## 7. Probabilistische koppeling met de statistiek van de doodsoorzaken

Terwijl het voor de gegevens van de volkstelling en het Rijksregister mogelijk was een koppeling te realiseren waarvan we zeker zijn dat de gekoppelde bestanden wel degelijk betrekking hebben op dezelfde persoon, vormt de koppeling met de data van het overlijdensregister een totaal ander probleem. De gegevens van het overlijdensregister zijn geanonimiseerd. Ze bevatten geen naam- of adresgegevens. Er is ook geen unieke identificatiesleutel (zoals het Rijksregisternummer) hetgeen de directe koppeling van bestanden onmogelijk maakt.

In tegenstelling met de *deterministische* koppeling tussen Rijksregistergegevens en de volkstellingsdata van 1991 is hier sprake van een *probabilistische* koppeling waarbij de gekoppelde gegevens met een zekere waarschijnlijkheid, maar niet met zekerheid, aan dezelfde persoon kunnen worden toegewezen.

Om een dergelijke probabilistische koppeling te realiseren zijn er een aantal voorwaarden:

1. uniciteit: er moeten voldoende variabelen aanwezig zijn die in combinatie met elkaar relatief uniek zijn. De variabelen moeten zodanig worden gekozen dat het aantal mogelijke duplicaten tot een minimum is herleid. Het spreekt vanzelf dat we de koppeling niet onnodig zwaar moeten maken door alle mogelijke variabelen te gebruiken, maar onze criteria dusdanig moeten kiezen dat met een minimale set gegevens een maximaal discriminerend vermogen kan worden gerealiseerd.

2. Een tweede voorwaarde is dat de variabelen die we willen gebruiken in beide te koppelen bestanden moeten voorkomen, anders is een koppeling sowieso uitgesloten. Deze voorwaarden leken ons hier voldoende vervuld om een koppeling te realiseren met het sterfteregister en zodoende socio-economische gegevens te kunnen verbinden met sterfteoorzaak. We beschikken immers in beide databanken over het geslacht, de

geboortedatum, de sterftedatum, de gemeente van verblijf en – in voorkomend geval – de huwelijksdatum.

Bij gebruik van deze variabelen krijgen we een combinatie die vrij uniek is voor elke persoon. Om een zicht te krijgen op de mate van uniciteit die kan worden bekomen via de verschillende combinaties, hebben we eerst die uniciteit getest binnen elke databank afzonderlijk (“within-file matching” of “internal linkage”). Dit bestaat in het vergelijken van elke record met alle andere records van dezelfde file. Deze techniek laat ook toe om “tweelingen of meerlingen” op te sporen en eventueel foutief dubbel geregistreerde gegevens te detecteren. De combinatie van gemeente, geslacht, geboortedatum en sterftedatum levert op een populatie van 10 miljoen personen nog nauwelijks dubbels op. Alleen in grotere gemeenten bestaat er een reële kans om twee personen van hetzelfde geslacht te vinden die op dezelfde dag geboren werden en overleden. In de overlijdensaangiften van 1994 vinden we slechts één enkel paar met dezelfde informatie op deze vier variabelen. In 1993 vinden we twee personen terug waarvoor een dubbel bestaat (dus 4 personen in het totaal) op 107.085 overlijdens. Voor 1992 hebben we 9 paren op 104.324 personen en voor 1991 zelfs 10 paren. Met grote waarschijnlijkheid gaat het hier voor de jaren 1991 en 1992 in de meerderheid der gevallen om een dubbele registratie. Gezien het om geanonimiseerde gegevens gaat kunnen we dit niet controleren, maar voor het jaar 1993 hadden we oorspronkelijk 7 identieke paren gevonden. We hebben die laten controleren op het NIS waar men de mogelijkheid bezat om de oorspronkelijke gegevens na te trekken en daaruit bleek dat het in 5 gevallen om een dubbele registratie ging! We kunnen in elk geval besluiten dat deze combinatie van gegevens een bijzonder hoge graad van uniciteit bezit en dat bijgevolg de koppeling tussen twee bestanden op deze basis voor bijna 100% correct zal zijn op voorwaarde dat de desbetreffende variabelen juist zijn geregistreerd in beide bestanden.

## 8. De koppelingsmethode

Om de koppeling tussen databanken te realiseren waarvoor geen unieke sleutel bestaat kan men in het algemeen twee technieken ontwikkelen indien men over een aantal variabelen beschikt die in beide databanken voorkomen.

Een bekende techniek bestaat erin de databank verschillende malen na elkaar te doorlopen waarbij een variabele als vangnet wordt gebruikt en men daarna welbepaalde variaties toelaat op andere variabelen (Newcombe, 1988). Bij het vergelijken van twee bestanden met persoonsnamen wordt bijvoorbeeld eerst de familienaam als vangnet gebruikt en laat men een welbepaalde variatie toe op de datum van overlijden. In een tweede fase wordt de datum van overlijden als vangnet gebruikt: men gaat alle records met zelfde overlijdensdatum weerhouden. Binnen die groep gaat men vervolgens een aantal variaties toelaten op de naam binnen bepaalde grenzen die men logisch en aanvaardbaar acht. Het voordeel van deze techniek is dat men grenzen kan vastleggen waarbinnen men fouten tolereert. Het nadeel is dat het zeer omslachtig is en moeilijk aanwendbaar bij zeer grote databanken waar men met een aanzienlijk overschot kan blijven zitten dat nog manueel dient te worden bewerkt.

De andere benadering vertrekt van de aanmaak van een nieuwe variabele die door de combinatie vrij uniek wordt. Het voordeel van deze aanpak is dat bij een koppeling de kans relatief groot is dat het inderdaad om dezelfde persoon gaat. Die kans wordt groter naarmate de populatie waarbinnen men moet zoeken kleiner is, de combinatie van variabelen uitgebreider en naarmate elk van die variabelen meer verschillende waarden kan aannemen. Een geboortjaar gaat bijvoorbeeld slecht discrimineren. Binnen een populatie geeft dit een honderdtal mogelijkheden met een typische verdeling die overeenstemt met de bevolkingspyramide. Een geboortemaand scoort nog slechter met slechts 12 mogelijkheden, weliswaar relatief gelijkmatig verdeeld. Een geboortedatum geeft reeds een vrij goede discriminatie met meer dan 36000 mogelijke combinaties, hoewel ook hier ongelijkmatig verdeeld. De combinatie van de geboortedatum met andere variabelen geeft dus snel een steeds grotere kans op een unieke combinatie. Het nadeel van deze aanpak is dat de gegevens absoluut correct moeten zijn. De geringste afwijking, bijvoorbeeld één dag verschil in de geboortedatum, geeft aanleiding tot een niet-koppeling (false negative match). Er kunnen natuurlijk ook verkeerde koppelingen (false positive match) ontstaan door foute gegevens. De kans op een verkeerde koppeling door foute gegevens wordt echter kleiner naarmate we meer variabelen combineren. Een complexe combinatie aan variabelen vermindert bijgevolg de kans op een onterechte koppeling zowel door fouten als door doublures.

Bij het aanmaken van omvangrijke bestanden kan er altijd een zeker foutenpercentage optreden in de registratie van de gegevens. De wijze waarop de sterfteregisters tot stand kwamen, met veel manuele tussenstappen, liet toe te vermoeden dat het foutenpercentage wel relatief hoog kon zijn. Na het doorlopen van de testprogramma's kwamen we tot de vaststelling dat ongeveer 15% van de records uit de sterfteoorzaken niet werden gekoppeld met de volkstellingsgegevens op basis van de vier gebruikte variabelen (geslacht, gemeente van verblijf, geboortedatum en sterftedatum). Een deel van die registraties zijn misschien niet te koppelen omdat ze hun tegenhanger niet hebben in de andere databank. Het merendeel van de niet-gekoppelde records was echter het gevolg van fouten in één van beide te koppelen databanken.

Het kwam er dus op aan om te proberen zoveel mogelijk van die restgroep te koppelen door de precisie, de uniciteit van de combinatie, te verlagen. Dit stelt het probleem van de verhouding tussen de kans om een correcte koppeling tot stand te brengen en het risico dat het om een foutieve koppeling gaat. Het spreekt vanzelf dat wanneer die kansen gelijk zijn een koppeling nog weinig zin heeft.

Ter illustratie geven we de gedetailleerde resultaten voor de sterfte in 1994. We hebben een databank met overliddens naar oorzaak met 104.160 personen. Volgens het Rijksregister waren er dit jaar 103.566 overliddens. Van de bevolking die in 1991 in de volkstelling werd geregistreerd vinden we 102.010 personen die in 1994 overlijden volgens de gegevens van het Rijksregister. Een aantal andere personen uit de volkstelling zijn eveneens overleden maar na migratie. Theoretisch hebben we een preciese registratie van alle personen uit de volkstelling die in België gebleven zijn en in 1994 zijn overleden. De bijkomende overliddens in het Rijksregister zijn het gevolg van sterfte van personen die na 1 maart 1991 geboren werden of immigreerden. Daarvan kunnen er

86.995 worden verbonden met de databank van overlijdens naar oorzaak. Er blijven dus 15.009 records over van personen uit de volkstelling van 1991 die volgens het Rijksregister overleden maar waarvoor geen overlijdenscertificaat wordt gevonden. Aan de andere kant hebben we een databank met 17.165 overlijdenscertificaten waarvoor geen match gevonden werd in het volkstellingsbestand.

Vanwaar die grote verschillen? Gezien de relatief unieke combinatie van gegevens zouden deze grotendeels moeten overeenstemmen. De niet gekoppelde records zijn te herleiden tot drie verschillende groepen:

1. Een eerste groep bestaat uit personen die na de volkstelling in België zijn komen wonen of na 1/3/91 geboren werden en die in de loop van 1992 overleden. Deze groep genereert niet te koppelen records in het bestand van de overlijdenscertificaten.

2. Een tweede groep bestaat uit personen die weliswaar in België geregistreerd zijn, maar die in het buitenland overlijden. Van hen wordt geen overlijdenscertificaat opgemaakt. Wellicht is dit een belangrijke groep. Hieronder vallen niet alleen Belgische toeristen in het buitenland, maar ook personen die studeren of werken in het buitenland maar hun verblijfplaats aanhouden in België. Onder deze groep vinden we bijvoorbeeld ook gepensioneerden die in Spanje verblijven of oudere migranten die lange periodes in hun land van herkomst doorbrengen (Italië, Spanje, Griekenland, Turkije,...) maar in België ingeschreven blijven. Omgekeerd zijn er dus ook overlijdenscertificaten die worden opgemaakt van personen die op het Belgisch grondgebied overlijden en niet in de bevolkingsregisters of vreemdelingenregisters zijn opgenomen. Voor het jaar 1992 tellen we in de databank met overlijdens van personen ouder dan 1 jaar 1.134 certificaten van personen die niet in België geregistreerd staan. Van deze records uit de overlijdenscertificaten vinden we uiteraard geen tegenhangers in het bestand van de volkstelling. We mogen aannemen dat de groep residenten die in het buitenland overlijden ongeveer een zelfde omvang kan hebben. De groep overleden personen uit de volkstelling waarvoor we geen certificaat kunnen vinden kan dus gemakkelijk een duizendtal personen tellen.

3. Tenslotte bestaat de restgroep uit records met fouten. Het volstaat inderdaad dat slechts een klein stukje informatie verschillend geregistreerd werd in het Rijksregister en in het overlijdenscertificaat opdat beide niet gekoppeld kunnen worden. Waar ligt de fout? Wellicht hebben we – wat de gegevens betreft die we voor onze koppeling gebruiken – twee soorten « fouten ». Enerzijds het niet overeenstemmen van de gemeentecodes en anderzijds een foutieve registratie van gemeente, geboortedatum, sterftedatum of geslacht. Waarom dit onderscheid? Omdat de informatie « gemeente » hier een bijzondere plaats inneemt. Wat de gemeente betreft hebben we niet alleen te maken met « foutieve » gegevens maar ook met tal van personen waarvan de feitelijke verblijfplaats niet overeenstemt met hun woonplaats in het Rijksregister. Dit kan zowel een persoon zijn die niet officieel is uitgeschreven als een gewone verschuiving in de tijd. Verhuisbewegingen worden niet direct opgetekend en/of aangegeven en het overlijden kan plaats vinden vooraleer dit is gebeurd. Het hoeft dan ook niet te verbazen dat het loslaten van het criterium « gemeente van verblijf » toelaat om de grootste groep van de « missing » records toe te wijzen. Met uitzondering van de gemeente zijn de fouten vermoedelijk voor meer dan 99% in de fiches te vinden en in de achtereenvolgende stadia van registratie die deze doorlopen. Vanaf de eerste stap, de schriftelijke registratie door de geneesheer die de doodsoorzaak vaststelt, worden onvermijdelijk al talrijke schrijf- of

leesfouten gegenereerd. De daaropvolgende transcripties zorgen voor een toename in de fouten die wellicht nauwelijks worden gecompenseerd door de verbeteringen die in de verschillende stadia van verwerking van de gegevens worden aangebracht.

De methode die we gebruikt hebben om die resterende records te koppelen is gebaseerd op het reduceren van de koppelingsleutel door er een stukje informatie uit weg te laten. Gezien uit verkennend onderzoek was gebleken dat heel wat foutieve records het gevolg van schrijffouten waren, opteerden we voor het opsplitsen van alle gegevens in zo klein mogelijke eenheden. Vaak blijkt in een datum alleen de maand fout te zijn genoteerd, of zijn maand en dag door elkaar gehaald. In plaats van alle informatie over de geboortedatum weg te laten, werd deze in drie afzonderlijke variabelen opgesplitst die elk afzonderlijk kunnen worden getoetst. In een eerste run op de restgroep werd eerst en vooral de gemeente weggelaten. Op die manier konden niet minder dan 8.954 records gekoppeld worden. Dit stemde overeen met onze assumptie dat de meeste fouten in deze variabele zouden kunnen opduiken. Het weglaten van de gemeente vergroot natuurlijk de kans op onterechte koppelingen.

Een goede graadmeter hiervoor is het toepassen van de techniek van de interne linkage. Alvorens de matching met de andere file door te voeren, werden telkens alle tweelingen (of meerlingen) uit elk bestand gehaald. Voor de gemeenten vonden we in het bestand 1994 op 15.009 records 28 personen met een dubbelganger van hetzelfde geslacht die op precies dezelfde datum geboren is en overleden. Het gaat hier om verschillende personen die op dit beperkt aantal variabelen gelijk scoren. Deze 28 paren werden weggelaten bij de koppeling tussen de twee bestanden en achteraf terug toegevoegd bij het te koppelen restbestand. Het weghalen van dubbels vermijdt het maken van valse koppelingen, gezien voor elke dubbel twee potentiële koppelingen mogelijk zijn. Het achteraf terug toevoegen aan de pool laat toe dat bij een andere combinatie een match toch nog wordt gerealiseerd. Het belang van dit proces is echter ook dat het meteen een aanduiding geeft over de omvang van mogelijke foute koppelingen. Een interne koppeling is in wezen een louter toevallige koppeling op een aantal variabelen die aan totaal verschillende personen toebehoren. Het aantal gevonden dubbels geeft dus een aanduiding van de kans om valse koppelingen te maken. Als er geen dubbels worden gevonden betekent dit niet dat een foute koppeling is uitgesloten, maar dat de combinatie van variabelen zodanig is dat binnen het gegeven bestand de kans op een foute koppeling relatief klein is. De onderlinge volgorde waarin de verschillende stappen werden uitgevoerd bleek geen invloed te hebben op voorwaarde dat wel een volgorde wordt gerespecteerd in het aantal elementen dat wordt weggelaten.

In tabel 2 geven we een overzicht van het volledige koppelingsproces en resultaten van de controle op dubbels voor de sterftcertificaten van 1994. De controle op de dubbels kan zowel gebeuren binnen de restgroep van de volkstelling als binnen de restgroep van sterftcertificaten. Wij deden het voor beide maar gebruikten de restgroep van de volkstelling als berekeningsbasis. Bij de sterftcertificaten bestaat inderdaad de mogelijkheid dat er naast de valse dubbels ook echte dubbels voorkomen (door een dubbele registratie).



We beperken de tabel tot de stappen die in een eerste analyse bruikbaar leken. In de testfase zijn we aanvankelijk veel verder gegaan in het aantal sleutelreducties en in het testen van de resultaatfiles.

Het grootste aantal fouten kan potentieel ontstaan bij het weglaten van de gemeente. Deze variabele kan 589 verschillende waarden aannemen. Het weglaten van de geboortedag daarentegen levert geen dubbels op (maximaal 31 verschillende waarden), evenals de geboortemaand (12 verschillende waarden). Het is slechts het geboortjaar (ongeveer 100 waarden) en de combinatie tussen geboortemaand en geboortedag (365 waarden) die opnieuw dubbels opleveren. Aan de hand van het aantal dubbels bij interne linkage kunnen we gerust stellen dat het aantal foutieve koppelingen zeer miniem is tegenover het aantal gerealiseerde koppelingen. Bij het weglaten van de gemeente hebben we minder dan een 0,4% kans op een foutieve koppeling terwijl we bijna 60% van de resterende records kunnen koppelen.

Reductie van de koppelingsleutel door	Restgroep uit census	dubbels	nieuwe koppelingen	% dubbels / nieuwe recs
Weglaten gemeente	15.009	28	8.954	0,31
Weglaten geboortedag	6.055	0	605	0
Weglaten geboortemaand	5.450	0	536	0
Weglaten geboortjaar	4.914	3	677	0,44
Weglaten geboortemaand en –dag	4.237	2	227	0,88
Weglaten dag van overlijden	4.010	0	1.239	0
Weglaten maand van overlijden	2.771	0	175	0
Weglaten maand en dag van overlijden	2.596	2	56	3,57
Weglaten geslacht	2.540	0	478	0
Weglaten gemeente en dag van overl.	2.062	27	176	15,3
Weglaten gemeente en geboortjaar	1.886	17	119	14,29
Weglaten gemeente en geboortedag	1.767	12	89	13,48
Weglaten gemeente en geboortemaand	1.678	7	63	11,11
Weglaten gemeente en maand van overl.	1.615	11	2	550
Weglaten geslacht en geboortedag	1.613	0	3	0
Weglaten geslacht en geboortemaand	1.610	0	3	0
Weglaten geslacht en geboortjaar	1.607	0	5	0
Weglaten gesl., geboortemaand en –dag	1.602	1	1	100
Weglaten geslacht en dag overlijden	1.601	0	33	0
Weglaten geslacht en maand overlijden	1.568	0	3	0
Weglaten geslacht en gemeente	1.565	7	57	12,28
Uitme restgroep uit census	1.508			

*Tabel 2: Overzicht koppelingsproces tussen sterfteregister 1994 enerzijds en de gegevens uit de volkstelling. De eerste kolom geeft de records die niet werden gekoppeld op basis van de volledige identificatiesleutel. Voor elke stap wordt de restgroep aangegeven en het aantal dubbels bij de aanmaak van een beperktere sleutel. De laatste kolom geeft de verhouding tussen het aantal bijkomende koppelingen en het aantal aanwezige dubbels*

We zien dat de verhouding merkkelijk minder gunstig wordt wanneer we de gemeente en een andere variabele samen weglaten. Het weglaten van de gemeente en de sterftedag geeft reeds 2,7% (27 op 2.062) potentiële fouten en levert slechts 8,5% (176 op 2.062) nieuwe koppelingen op. Bij het weglaten van gemeente en sterftemaand is het aantal potentiële fouten zelfs aanzienlijk groter dan de nog gerealiseerde koppelingen. Bijkomende gekoppelde records gaan hier ten koste van een nog grotere toename in mogelijke foutieve koppelingen. In de definitieve versie van het programma zullen we dan ook de stappen waar het aantal potentiële fouten het aantal nieuw gerealiseerde koppelingen evenaart of overtreft weglaten. Dit is het geval voor de combinaties waar gemeente en sterftemaand worden weggelaten en waar geslacht, geboortemaand en geboortedag worden weggelaten. In alle overige stappen is het aantal potentiële fouten aanzienlijk lager dan de gerealiseerde winst.

In deze tabel zijn overigens twee merkwaaardige vaststellingen te noteren. Ten eerste blijken er aanzienlijk wat geslachtsomkeringen op te treden. Voor 1994 hebben we niet minder dan 478 personen die we kunnen koppelen bij eenvoudige weglating van het geslacht. Bij combinatie van weglating van het geslacht en een andere variabele komen daar nog meer dan 100 personen bij. Dit stelt meteen ook het probleem van de correctiemethode die tot op heden op deze bestanden werd toegepast. Eén van de consistentiechecks die men op het NIS gebruikte was het corrigeren van sterfteoorzaken die geslachtsincompatibel waren. Het teruggrijpen naar het Rijksregister toont echter dat voor ruim 0,5% van de sterftecertificaten het geslacht verkeerd werd gecodeerd.

Ook het hoog aantal fouten bij de sterftedatum - en meer in het bijzonder de dag van overlijden - is merkwaardig. Dit is immers de datum die direct met de gebeurtenis zelf van het overlijden is gekoppeld en bijgevolg in elk geval bijna steeds zeer accuraat kan worden vastgesteld. De mogelijkheid werd geopperd dat problemen van fiscaliteit of successierechten hier een rol zouden kunnen spelen. Dit lijkt ons niet het geval. Bij de afwijking tussen de datum in het Rijksregister en het sterftecertificaat is geen systematische fout te ontdekken. De afwijkingen zijn normaal verdeeld zonder merkkelijk hogere frequentie voor of na de in het register opgetekende datum van overlijden. Vijftig procent van de fouten zijn slechts één dag vroeger of later. Een mogelijke verklaring zou kunnen liggen in het feit dat de geboortedatum door de gemeenteambtenaar in het register kan worden gecontroleerd en dat bij een gelijk aantal foutieve transcripties men op de geboortedatum gewoon veel meer correcties kan aanbrengen dan op de sterftedatum die nog niet in het register werd ingebracht. De cijfers voor de jaren '91, '92 en '93 lopen behoorlijk parallel met deze resultaten. Daarenboven geeft ook de interne koppeling van de restgroep van de sterftecertificaten een aanduiding van het aantal mogelijke valse koppelingen, hoewel we hier misschien ook te maken hebben met dubbele inschrijvingen. Het feit dat het aantal dubbels systematisch nogal wat hoger ligt dan bij de restgroep van de census zou hier kunnen op duiden. Hoe dan ook gaat het nog steeds om zeer kleine aantallen.

Voor de koppeling van de bestanden werd het programma uiteindelijk herschreven en werden alle modules weggelaten indien bij één van de vier onderzochte jaren het aantal interne dubbels hoger of gelijk was aan het aantal nieuw gekoppelde bestanden. De

verschillende jaren waren in dit opzicht onderling zeer consistent. De modules die het risico bevatten op een groter aantal foutieve dan correcte koppelingen bleken over alle jaren heen gelijk te zijn en beperkt tot de twee hierboven geciteerde.

## 9. Gebruik van controlevariabelen

Een bijkomende methode om te controleren in welke mate onze probabilistische koppeling goede resultaten boekt is het gebruik van controlevariabelen. In beide te koppelen bestanden beschikken we ook nog over de huwelijksdatum. Het probleem met deze variabele is dat hij niet voor iedereen van toepassing is. Als koppelingsvariabele is hij dus slecht te gebruiken. Daarenboven ontbreken deze gegevens op heel wat sterftecertificaten. Controle op de huwelijksdatum kan dus geen uitsluitel geven over het aantal juiste koppelingen, het is wel een goede graadmeter. De huwelijksdatum kan alleen een afnemende of toenemende kans op fout geven en dit hebben we vertaald in een nieuwe controlevariabele. Deze variabelen kan volgende waarden aannemen:

*Code 0* : de huwelijksdatum is ingevuld en gelijk in beide bestanden. Voor deze records hebben we vrijwel 100% zekerheid dat het om dezelfde persoon gaat. Hier zou men nog een stap verder kunnen gaan door ook gebruik te maken van de geboortedatum van de partner. Gezien deze echter in de meeste certificaten ontbreekt en het gebruik ervan weinig toevoegt aan de controle die we kunnen uitvoeren met de huwelijksdatum, laten we dit gemakshalve achterwege.

*Code 1* : de huwelijksdatum is verschillend in beide bestanden. Theoretisch zou dit hier om een foute koppeling moeten gaan. Het is echter mogelijk dat de huwelijksdatum verkeerd werd ingevuld. Gezien de veelvuldige fouten in de geboortedatum en de sterftedatum gaan we er van uit dat dit in een aantal (en misschien zelfs in de meeste) gevallen inderdaad zo is. Maar we kunnen stellen dat naarmate we hier een groter percentage hebben, de kans groeit dat we personen gekoppeld hebben die niet bij elkaar horen.

*Codes 2 tot 4* : voor al deze waarden geldt dat de koppeling correct kan zijn met een foutenkans die ergens tussen code 0 en code 1 ligt.

*Code 4* : personen die niet gehuwd zijn. Alleen het al dan niet gehuwd zijn vormt een controlefactor.

*Code 3* : Er is geen huwelijksdatum op het sterftecertificaat ingevuld hoewel er wel een huwelijksdatum ingevuld is in het Rijksregister. Vermoedelijk gaat het om een nalatigheid bij het invullen van het sterftecertificaat. Het geeft ons geen enkele controle op de juistheid van de koppeling.

*Code 2* : huwelijksdatum als onbekend genoteerd op het certificaat en er is een huwelijksdatum vermeld in het Rijksregister. Dit geeft ons eveneens geen enkele controlemogelijkheid.

Wanneer we deze controle uitvoeren op de diverse stadia van het koppelingsproces geeft ons dit wel een mogelijke graadmeter over de betrouwbaarheid van onze koppelingen. We dienen inderdaad rekening te houden met het feit dat ook de fouten niet willekeurig over de records verdeeld zijn. Het zou best kunnen dat records met fouten in de gemeente van verblijf ook een grotere foutkans hebben op alle andere variabelen met inbegrip van de huwelijksdatum.

Wanneer we opnieuw de koppeling voor het jaar '94 als basis nemen blijkt dat bij de basiskoppeling 42% van de 86.995 records exact hetzelfde huwelijksjaar aangeven en 1,6% een verschillend jaar aangeven (code 1). We kunnen dit ook controleren op huwelijksmaand en huwelijksdag; deze geven respectievelijk de waarden 42% en 1,5% voor de huwelijksmaand en 41,4% en 2,1% voor de huwelijksdag. Deze waarden overlappen elkaar gedeeltelijk maar niet volledig.

Gezien de eerste match werd doorgevoerd op personen die in dezelfde gemeente zijn ingeschreven, hetzelfde geslacht hebben en op dezelfde datum geboren zijn en overleden mogen we aannemen dat de koppeling hier vrijwel volledig correct is. Ter herinnering: voor 1994 zijn er voor het gehele land slechts 2 personen op exact dezelfde datum overleden die van hetzelfde geslacht waren, geboren op dezelfde datum en woonachtig in dezelfde gemeente. Zelfs wanneer we het criterium van de gemeente laten vallen zijn minder dan 0,4% personen van hetzelfde geslacht op dezelfde dag geboren en overleden. Dit betekent dat 1,6% van de huwelijksdata die niet overeenstemmen wat het jaar betreft tussen het Rijksregister en de fiches vermoedelijk een goede maat zijn voor het aantal fout geregistreerde huwelijksdata. Onze verdere koppelingen met een minder preciese sleutel moeten dus zo dicht mogelijk bij die 1,6% liggen om betrouwbaar te zijn. Omgekeerd halen we 42% van de paren die exact matchen in het basisbestand. Hoe dichter we bij die 42% geraken, hoe betrouwbaarder onze koppeling.

Koppelingsprocedure	% dat gelijk is en % dat verschillend is voor		
	Jaar	Maand	Dag
Koppeling zonder gemeente	24/2,5	24/2,4	23,7/2,8
Koppeling zonder geboortedag	37,2/2	37,2/2	36,9/2,3
Koppeling zonder geboortemaand	38,8/2,8	39,6/2,1	38,1/3,5
Koppeling zonder geboortejaar	35,9/2,5	36,5/1,8	34,9/3,4
Koppeling zonder geb. maand & dag	49,8/4,4	37,4/5,3	37/5,7
Koppeling zonder dag van overlijden	35,2/1,5	35,4/1,2	34,8/1,9
Koppeling zonder maand van overlijden	37,1/1,1	36,9/2,3	34,9/3,4
Koppeling zonder maand & dag overlijden	42,9/3,6	42,9/3,6	41,1/5,4

*Tabel 3: Per koppelingsprocedure percentage huwelijksdata dat identiek is en percentage dat verschillend is tussen de gekoppelde records voor elk onderdeel van de huwelijksdatum*

In tabel 3 worden de resultaten geïllustreerd voor de eerste stadia van het koppelingsproces voor de gegevens van 1994. Daaruit blijkt dat mogelijke fouten bij het weglaten van de gemeente niet veel hoger liggen dan bij de koppeling met volledige sleutel (van 2,4 tot 2,8% tegenover 1,5 tot 2,1%) en eerder lager dan bij alle andere weglatingen waar de potentiële fouten kunnen oplopen van 4,4 tot 5,7%. We zien overigens dat bij het weglaten van twee elementen het aantal potentiële fouten het hoogst oploopt. Het beeld van het aantal koppelingen die met grote waarschijnlijkheid correct zijn verloopt enigszins anders. De scores blijken voor alle weglatingen nog redelijk hoog te liggen (van 35% tot niet minder dan 50%) met uitzondering van de koppeling met

weglating van de gemeentecode. Vermoedelijk ligt de verklaring hiervoor in het feit dat precies het ontbreken van een correcte codering van de gemeente van verblijf ook de kans vergroot op een gebrekkig invullen van de andere administratieve gegevens en dat ook de controle door de gemeente achteraf hier meer kans loopt om in gebreke te blijven. Over het algemeen mogen we besluiten dat de koppeling met weglating van bovenstaande variabelen zeker nog verantwoord is en over het algemeen onze databank aanvult met aanzienlijk meer correcte koppelingen dan potentieel foute koppelingen.

Tot nu toe zijn we er in feite impliciet van uitgegaan dat het jaar van overlijden een correct gegeven is. We gingen uit van de bedenking dat gezien de bestanden van de sterftcertificaten per jaar worden aangemaakt dat deze op dit punt correct waren. Praktisch voordeel van deze assumptie was dat we het programma op een afzonderlijk jaar konden laten lopen wat handelbaarheid en overzichtelijkheid ten goede kwam. Na het draaien van het koppelingsprogramma per jaar was het logisch dat de vraag zich stelde of er ook geen vergissing mogelijk was in het jaar van overlijden. Voor de vier jaar samen hadden we 6.596 personen uit de volkstelling waarvoor we geen doodsoorzaak gevonden hadden. Anderzijds hadden we nog 13.179 ongekoppelde records uit het sterfteregister. Het grootste deel van deze restfiles bestaat uit records die niet te koppelen zijn omdat ze andere groepen bestrijken (zie supra). Maar omdat misschien nog een beperkt aantal in een verkeerd jaar waren terecht gekomen, werd het programma nog eens gedraaid op de twee restfiles. In feite betekent dit het weglaten van een bijkomende controlevariabele - het sterftejaar - en dit voor elke koppelingsmodule. Het programma gaf 636 bijkomende gekoppelde records. Bij controle op dubbels binnen elk stap van het proces en bij gebruik van de controlevariabele bleek echter dat het aantal foutieve koppelingen te grote proporties nam. Het laten vallen van de geboortedag gaf bijvoorbeeld 50 gekoppelde bestanden. Controle op de huwelijksdatum leverde hier echter een volkomen andere verhouding op dan bij onze bestanden per jaar. Slechts 2% van de huwelijksdata kwamen overeen en 20% waren verschillend. Het grootste aantal koppelingen (208 van de 636) kwam tot stand bij het weglaten van de gemeente, het geboortjaar en het jaar van overlijden. Samen zijn dit echter de drie variabelen die het meest verschillende waarden kunnen aannemen en dus het grootste discriminerend vermogen hebben. Bij het controleren van duplicaten binnen het bestand van de volkstelling gaf het laten vallen van deze variabelen niet minder dan 118 paren of 236 records op een resterend totaal van 6.347 en dus meer kans op foute koppelingen dan het totaal dat we realiseerden. Toch bleek de controlevariabele van de huwelijksdatum hier 28,8% overeenstemmende huwelijksdata te geven. Een mogelijke optie zou dus kunnen zijn om de koppelingsprocedure door te voeren met weglating van het jaar van overlijden en van deze verwerking enkel de gekoppelde bestanden te behouden waar de huwelijksdatum overéénstemt.

We deden dit voorlopig niet omwille van het groot verschil in kwaliteit tussen de verschillende jaren van het sterfteregister. Voorlopig verkozen we elk jaar van het sterfteregister afzonderlijk te houden.

## 10. Finaal bestand

De koppeling met het sterfteregister geeft uiteindelijk volgend resultaat:

In het totaal werden 9.978.681 mensen geregistreerd in de volkstelling. Voor 27 werden geen gegevens teruggevonden in het Rijksregister. Van de overige 9.978.654 bleken er nog 9.203.621 in België te wonen, 611.597 waren overleden, 51.805 geschrapt uit de bevolkingsregisters en nog eens 111.631 officieel geëmigreerd. 97 personen blijken reeds voor de volkstelling overleden te zijn. Ofwel gaat het hier om tellingsdocumenten die voortijdig of onterecht werden teruggestuurd, ofwel gaat het hier om een foutieve overlijdensdatum. Ook 406 schrappingen en 65 emigraties bleken in 1991 vóór de volkstellingsdag te hebben plaats gevonden. Al deze personen werden uit het bestand geschrapt.

Jaar	overlijdens	schrappingen	Emigraties
91	84.663	6.737	21.026
92	102.489	7.762	20.426
93	105.181	7.464	18.020
94	102.010	7.874	17.262
95	103.115	9.071	16.294
96	20.332	1.598	2.537
Totaal	517.790	40.506	95.565

*Tabel 4: Overzichtstabel met aantal personen dat jaarlijks uit het censusbestand verdwijnt in de periode van 5 jaar volgend op de volkstelling van 1 maart 1991*

Gezien de sterfte wordt onderzocht tot 5 jaar na de volkstelling (1 maart 1996), wordt geen rekening gehouden met het verdwijnen uit de populatie na die datum. Het basisbestand omvat dus een totale populatie van 9.978.086 personen waarvan er 517.790 overlijden binnen de 5 jaar en nog eens 136.071 uit de populatie verdwijnen door emigratie. De verdeling over de verschillende jaren staat in tabel 4.

Jaar	N in Volkstelling en Rijksregister	N in sterfteregister	N gekoppeld	% gekoppeld op basis van VT
91	84.663	104.137	82.975	98%
92	102.489	104.306	100.697	98,3%
93	105.181	107.081	103.500	98,4%
94	102.010	104.160	100.493	98,5%
Totaal	394.343	419.684	387.665	98,3%

*Tabel 5: Aantal records met gekoppelde informatie uit sterfteregister en volkstelling*

De resultaten van de koppeling met het sterfteregister voor de sterfgevallen van de jaren '91 tot '94 staan in tabel 5. Het spreekt vanzelf dat bij het beschikbaar komen van de sterftecertificaten voor de jaren 1995 en 1996 we deze databank verder kunnen uitbreiden.

*Dit betekent dat, voor de personen die behoorden tot de volkstelling, we in staat waren om voor meer dan 98% van de overlijdens in de periode '91-94 de sterfteoorzaak op te sporen.* Toch bevat het bestand wellicht ook een klein aantal foutieve koppelingen en is het dus voor een individueel geval niet vanzelfsprekend om te stellen dat de socio-economische gegevens en de sterfteoorzaak wel degelijk betrekking hebben op dezelfde persoon. Door het groot aantal cases en het feit dat we met stelligheid weten dat, voor het leeuwendeel van de koppelingen, het om dezelfde persoon gaat, is deze fout echter verwaarloosbaar. Voor het geheel van de databank geeft de meest conservatieve berekening een maximum van 0,2% records die fout kunnen gekoppeld zijn naar sterfteoorzaak.

## 11. Overzicht van de covariaten

Men kan stellen dat sterfte wordt bepaald door drie groepen determinanten: biologische, structurele en culturele of gedragsdeterminanten. De variabelen van de volkstelling behoren bijna allen tot de structurele determinanten van sterfte. Alleen geslacht en leeftijd behoren tot de biologische determinanten. De volkstelling geeft verder geen enkele informatie over culturele determinanten zoals eetgewoonten, rookgewoonten, doktersbezoek, enz. In dit opzicht zou het heel verhelderend zijn mochten we de resultaten van het gezondheidsonderzoek, dat in hoofdzaak dergelijke informatie bevat, kunnen verbinden met de nationale databank mortaliteit.

De structurele determinanten kunnen op hun beurt worden ingedeeld in drie grote groepen:

1. De invloed die wordt uitgeoefend door het *milieu*, d.i. door de woon- en werkomgeving, op de gezondheid van het individu. Een hele reeks variabelen kunnen ons hieromtrent informatie verschaffen:

- Woonplaats + duur van verblijf
- Plaats van tewerkstelling
- Sector van tewerkstelling
- Woonomstandigheden (comfort)
- Woon- werkverplaatsingen

2. *De sociaal-economische status.* Wat is de invloed op gezondheid en op sterfterisico van de gevolgde opleiding, van de maatschappelijke status en van het verworven inkomen? Of omgekeerd wat is het effect van de toegang tot informatie (onderwijs) en de toegang tot schaarse goederen (inkomen) waaronder gezondheidszorg? Klassiek worden drie indicatoren voor de sociaal-economische status (SES) opgemeten: opleidingsniveau, beroep en inkomen.

Sociaal-economische status kan natuurlijk ook aanleiding geven tot bepaalde patronen van consumptie- of gezondheidsgedrag die als het ware de transmissieriem vormen die de invloed van SES op sterfterisico verklaren. Maar SES bepaalt ook de woon- en de werkplaats met de bijhorende milieufactoren waarin men terecht komt. Het bepaalt tevens aan welke werkomstandigheden men wordt onderworpen (ritme, fysieke inspanningen, stress,...). Voor onderwijs en tewerkstelling vinden we in de volkstelling vrij goede informatie. Voor inkomen kunnen we alleen werken met proxies. Het woningcomfort in combinatie met het bewonersstatuut (eigenaar, huurder) geeft – op enkele uitzonderingen na - een zeer goede schatting van het inkomen. Laag wooncomfort duidt over het algemeen op een laag inkomen. Wie een voldoende hoog inkomen heeft zal niet lang in slechte woonomstandigheden blijven wonen. Omgekeerd kunnen lage inkomens niet lang de last dragen van te hoge uitgaven in de woonkosten. Daarenboven kent de woningvariabele een zekere inertie. Het duurt een tijdje voor inkomensdaling of –stijging zich vertaalt in een andere woning. Het is dus een vrij stabiele variabele. Voor leeftijdsgroepen boven de 30 jaar mogen we aannemen dat het vaak om een betere vertaling gaat van inkomen dan sommige fiscale aangiften.

In de volkstellingsgegevens vinden de drie SES-indicatoren hun vertaling in volgende covariaten:

Onderwijs

- Hoogst behaalde diploma
- Leeftijd bij het beëindigen van de studies
- Hoogst behaalde diploma in het huishouden

Tewerkstelling

- Personencategorie (scholier, werkloze, werkend,...)
- Statuut beroep (arbeider, bediende,...)
- Beroep
- Aantal uren tewerkstelling
- Beroep van de echtgenoot

Inkomen

- Type inkomen (uit arbeid, vervangingsinkomen, pensioen,...)
- Inkomensypes binnen het huishouden
- Huurder of eigenaar van woning
- Comfort van de woning
- Alle overige woningvariabelen (oppervlakte, aantal kamers, bouwjaar...)
- Wagenbezit

3. Een laatste groep van covariaten wordt gevormd door de *leefvorm*. Deze is niet zonder belang voor tal van gezondheidsindicatoren en gecorreleerd met de levensverwachting. In de volkstellingen geven een aantal variabelen hierover informatie:

- Burgerlijke stand
- Gezinsamenstelling
- Aantal leden binnen het huishouden
- Een aantal typologieën van huishoudens, waaronder de individuele “Lipro” huishoudensposities (Deboosere, 1997).



## 12. Besluit

1. Het onderzoek dat tot nu toe werd uitgevoerd geeft aan dat de aangemaakte databank van zeer groot nut kan zijn voor een beter begrip van de differentiële mortaliteit in ons land. *We weten nu ook dat het technisch mogelijk is deze databank verder aan te maken op permanente basis.* In dit opzicht zou het van groot belang zijn om een toekomstige volkstelling te kunnen linken aan de bestaande databank. Een nieuwe volkstelling zou het immers mogelijk maken om ook rekening te houden met de evolutie van de covariaten in de tijd. Voor sommige variabelen is dat geen groot probleem. Het onderwijsniveau dat men bereikt heeft blijft doorgaans gehandhaafd voor de duur van een mensenleven. Beroep en tewerkstelling echter kunnen in de loop van de tijd wel sterk evolueren. Een nieuwe volkstelling zou bovendien kunnen worden gebruikt om een minimaal aantal essentiële gegevens over gezondheidsgedrag te verzamelen.

2. *De uitwerking van deze databank heeft ook aangetoond dat meer dan 98% van de overlijdens van personen aanwezig in de volkstelling kan worden verbonden met een sterftecertificaat.* Dit is een bijna volledige koppeling gezien er voor het overlijden in het buitenland geen certificaat wordt opgemaakt.

3. *Het doorvoeren van een geautomatiseerde gegevensinvoer voor de sterftecertificaten zou de accuraatheid van de registratie aanzienlijk verbeteren met een veel eenvoudiger linkageprocedure als resultaat.* Het belangrijkste besluit is echter dat de linkageprocedure een krachtig hulpmiddel kan zijn om de kwaliteit van het sterfteregister – althans wat de administratieve data betreft – sterk te verbeteren. Op basis van onze analyse is gebleken dat administratieve fouten voor minstens 90% zouden kunnen worden gecorrigeerd. Het feit dat hierbij ook een aantal onterechte correcties kunnen optreden weegt niet op tegen de aanzienlijke kwaliteitsverbetering. Tenslotte gaat het in de huidige bestanden om niet minder dan 15% van de records die minstens één fout bevatten in de administratieve gegevens.

4. *Het jaarlijks aanmaken van de nationale databank mortaliteit kan aldus zowel een doel op zich zijn als een belangrijk hulpmiddel in het corrigeren van de sterftecertificaten.*

## Bibliografie

Aelvoet, W., F. Capet, J. Vanoverloop (1998), Gezondheidsindicatoren 1996, Ministerie van de Vlaamse Gemeenschap.

Aelvoet, W., M. Fortuin, P. Hooft, J. Vanoverloop (1999), Gezondheidsindicatoren 1997, Ministerie van de Vlaamse Gemeenschap.

Census Belgica 2001, (1997) Verslagboek deel 2, Tweedaagse studiedag over het gebruik en de toekomst van de Volkstelling in België, Steunpunt WAV, Steunpunt voor Demografie, DWTC, Leuven, pp. 116.

Deboosere, P., R. Lesthaeghe, J. Surkyn, P.-M. Boulanger, A. Lambert (1997), *Huishoudens en gezinnen*, Algemene Volks- en Woningtelling, Monografie n°4, Brussel, Nationaal Instituut voor de Statistiek, pp. 224.

Delanghe, L. (1971), *Differentiële sterfte in BELGIË. Een sociaal-demografische analyse*, Leuven, KUL, pp. 232.

Kunst, A.E. en J.P. Mackenbach (1989), "Sociale ongelijkheid en sterfte. Worden sociaal-economische verschillen in sterfte kleiner?", *Demos*, 6(6), pp. 47-48.

Kunst, A.E. (1997), *Cross-national comparisons of socio-economic differences in mortality*, Department of Public Health, Den Haag, Erasmus Universiteit Rotterdam, pp. 264.

Lagasse, R., I. Godin, et al. (1993), "De eerste doelstelling van de wereldgezondheidsorganisatie: gelijkheid op vlak van gezondheid. Een status quaestionis voor België", in *Sociale ongelijkheid en verschillen in gezondheid*, V. Raes, E. Kerkhofs en F. Louckx, Brussels, VUBPRESS, pp. 75-90.

Lagasse, R., P.C. Humblet, et al. (1990), "Health and social inequities in Belgium", *Social Science and Medicine*, 31(3), pp. 237-248.

Nationaal Instituut voor de Statistiek (1991), *Onderrichtingen aan de gemeentebesturen en Handleiding van de teller*, Brussel, pp. 51.

Newcombe, H.B. (1988), *Handbook of Record Linkage. Methods for health and statistical studies, administration, and business*, Oxford University Press, Oxford, pp.206.

Stoop, R., J. Surkyn (1997), In de wetenschap het niet te weten. Achtergronden van de non-respons in de Volkstelling van 1991, *Steunpunt voor Demografie Working Paper 1997-1*, Vrije Universiteit Brussel.

Van Oyen, H., J. Tafforeau, S. Drieskens, W. Aelvoet, S. Van den Broucke (1997), *Overlijden en oorzaaksspecifiek overlijden in de Vlaamse Gemeenschap 1989-1991*, Brussel, COOV, WIV.

Wanner, P., (1999), *Mortalité différentielle selon la cause de décès. Influence de la trajectoire de vie sur le risque de décès en Norvège (1960-1985)*, Louvain-la-Neuve, Academia, pp. 278.

Bijlage:

model overlijdenscertificaat zoals in gebruik tot 1998

**AANGIFTE VAN OVERLIJDEN VAN EEN PERSOON VAN EEN JAAR EN OUDER (a)**

**MODEL III C**

**STROOK A**

(strook in te vullen door de geneesheer en te bewaren door het gemeentebestuur)

Naam en voornaam van de overledene : .....

Echtgeno(o)t(e) of weduwe(naar) van : .....

Gewone verblijfplaats (gemeente, straat, nummer) : .....

Adres van overlijden : .....

Datum en uur van overlijden : .....

Handtekening en stempel van de geneesheer

N<sup>o</sup> van de overlijdensakte : .....

Ik bevestig, dat het overlijden te wijten is aan  
een natuurlijke, een geweldadige of een verdachte oorzaak

**STROOK B**

**AANGIFTE VAN OVERLIJDEN VAN EEN PERSOON VAN EEN JAAR EN OUDER (a)**

(strook in te vullen door de geneesheer, te verifiëren door het gemeentebestuur  
cf. strook D, en over te maken aan de gezondheidsinspecteur van de provincie)

1. Overlijdensdatum (d.m.j.) :

2. Plaats van overlijden :

- thuis  1
- ziekenhuis  2
- bejaardentehuis  3
- openbare weg  4
- werkplaats  5
- ambulantiewagen  6
- andere  7
- onbekend  9

3. Geslacht van de overledene : mannelijk  1

vrouwelijk  2

Vorbehouden aan de gezondheidsinspectie

- (a)
- (b) of E (b)
- (d)

De vakjes  moeten worden ingevuld met een cijfer, de vakjes  moeten worden ingevuld met een X

(a) Voor elk overlijden na de eerste verjaardag moet een telkaart worden ingevuld (voor de kinderen overleden vóór de leeftijd van één jaar wordt een "Aangifte van een doodgeboren kind of van het overlijden van een kind jonger dan één jaar" ingevuld.)

**STROOK C**

(strook in te vullen en te sluiten door de geneesheer)

Inlichtingen betreffende het overlijden :

Door ziekte

- (a) onmiddellijke oorzaak (vb. : broncho-pneumonie) : .....
- (b) oorspronkelijke oorzaak (vb. : mazelen) : .....

Door geweldadige  
oorzaak of toxische  
of farmaceutische  
stof

- (a) aard van trauma of stof : .....
- (vb. : schedelbreuk, barbituraten)
- E (b) feit of middel : .....
- (vb. : val van een trap, anafylactische reactie)

- gaat het om een ongeval  1
- een doding  2
- een zelfmoord  3
- onbekend  9

Datum, handtekening en stempel van de geneesheer

Bij verkeersongeval op de openbare weg, is het overlijden ingetreden :

- op het ogenblik van het ongeval  1
- binnen de 30 dagen  2
- na 30 dagen  3 (d)
- onbekend  9

Hier plooien

Hier plooien

N.B. de vakjes  moeten worden ingevuld met een cijfer

de vakjes  moeten worden ingevuld met een X

de vakjes  zijn voorbehouden aan het N.I.S.

**AANGIFTE VAN OVERLIJDEN VAN EEN PERSOON VAN EEN JAAR EN OUDER**

STROOK D

(strook in te vullen door het gemeentebestuur)

- 1. Gemeente (of land) van overlijden :
- 2. Nr van de overlijdensakte : .....
- 3. Gewone verblijfplaats van de overledene :
- Gemeente (of land) : .....
- 4. Geboortedatum (d.m.j.) :
- 5. Burgerlijke staat :
  - ongehuwd  1
  - gehuwd  2
  - weduwstaat  3
  - uit de echt gescheiden  4
  - gescheiden van tafel en bed  5
  - onbekend  9
- 6. Land van nationaliteit van de overledene :
  - België  00
  - ander (te preciseren)

- 7. Enkel indien de overledene gehuwd is :
  - Geboortedatum van de overlevende echtgeno(o)t(e) (d.m.j.) :
  - Datum van het laatste huwelijk (d.m.j.) :
- 8. Beroep of toestand (bv. student, werkloos, gepensioneerd, zonder beroep) :
- Eventueel, laatst uitgeoefend beroep :
- 9. Sociale beroepsgroep overeenstemmend naargelang van het geval, met het huidige of het laatst uitgeoefende beroep.
  - werkgever  1
  - zelfstandige  2
  - bediende  3
  - arbeider(ster)  4
  - helper(ster)  5
  - zonder  6
  - onbekend  9

Nr. van de overlijdensakte : .....

Gemeentestempel

Het gestolen vak mag, onder geen voorwendsel worden geopend, noch door de aangever, noch door om het even wie van het gemeentebestuur.

**BELANGRIJK**