

**RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEUHYGIENE
BILTHOVEN**

Rapport nr. 958606 001

**Overlevingsduuranalyse en concurrerende
doodsoorzaken. Een inleiding.**

R.T. Hoogmanveen, J.C. Jager

juni 1990

**Dit onderzoek werd verricht in opdracht en ten laste
van DGVCZ / Stafbureau Beleidsontwikkeling /
Stuurgroep Toekomstscenario's Gezondheidszorg.**

VERZAMELIJST

1-4	Stafbureau Beleidsontwikkeling van het Ministerie van WVC
5-7	Afdeling Epidemiologie en Informatica van het Ministerie van WVC
8	Directeur-Generaal van de Volksgezondheid
9	Plaatsvervangend Directeur-Generaal van de Volksgezondheid, tevens Hoofddirecteur Financiering en Planning
10	Drs. J.K.S. van Ginneken (NIDI)
11	Dr. L. Gunning-Schepers (Stabo, WVC)
12	Drs. S.E. Kooiker (NIVEL)
13	Dr. K. Schaapveld (NIPG/TNO)
14	Prof. drs. dr. ir. O.J. Vrieze (RL)
15	Depot Nederlandse publikaties en Nederlandse bibliografie
16	Directie Rijksinstituut voor Volksgezondheid en Milieuhygiëne
17	Ir. E.J.M. Feskens
18	Drs. S.H. Haisterkamp
19	Drs. J.A. Hoekstra
20	Prof.Dr.Ir. D. Kromhout
21	Drs. M.J.J.C. Poos
22	Drs. D. Ruwaard
23	Dr. W. Slob
24	N.G.W.M. van Strijp-Lockefeer
25	Dr. H. Verkleij
26-27	Auteurs
28	Bureau Projecten- en Rapportenregistratie
29-30	Bibliotheek RIVM
31-45	Reserve-exemplaren

VOORWOORD

Sinds 1983 laat de Stuurgroep Toekomstscenario's Gezondheidszorg (STG) toekomstverkenningen verrichten op het gebied van de volksgezondheid en gezondheidszorg. Hiervoor wordt meestal de zogenaamde scenario-methodiek gebruikt. Naarmate het aantal projecten toenam en de eerste scenario-rapporten beschikbaar kwamen, werd duidelijk, dat gezocht moest worden naar methoden om de (resultaten van de) verschillende scenario-onderzoeken te kunnen vergelijken en te combineren. Dit heeft geresulteerd in het project "Advisering en methodologie ontwikkeling STG projecten" (het methodologie-project), dat toegewezen is aan het Centrum voor Wiskundige Methoden (CWM) van het Rijksinstituut voor Volksgezondheid en Milieuhygiëne (RIVM).

Eén van de onderwerpen, die in het kader van het genoemde methodologie-project naar voren kwamen als een mogelijkheid om de resultaten van de verschillende scenario-onderzoeken te combineren, is het concept concurrerende doodsoorzaken. Hiermee wordt bedoeld, dat doodsoorzaken ten opzichte van elkaar concurrerend zijn. Bijvoorbeeld: preventie-programma's, die (dienen te) leiden tot een verandering van de sterfte aan hart- en vaatziekten, leiden eveneens tot een verandering van de sterfte aan kanker.

Het concept concurrerende doodsoorzaken speelt eveneens een rol bij de onderbouwing van gezondheidsdoelen voor toekomstig beleid, zoals beschreven in het Ontwerp-Kerndocument Gezondheidsbeleid. Deze gezondheidsdoelen hebben betrekking op het verlagen van (specifieke) mortaliteitscijfers naar een aantal doodsoorzaken, waarvoor de Wereld Gezondheids Organisatie (WHO) expliciete streefcijfers geformuleerd heeft. Het bovengenoemde voorbeeld geeft al aan, dat deze doelen met elkaar samenhangen. De Afdeling voor Epidemiologie en Informatica van het ministerie van WVC heeft aan het CWM en het Centrum voor Epidemiologie van het RIVM gevraagd om een voorstudie te verrichten op het gebied van concurrerende doodsoorzaken in het kader van het genoemde Ontwerp-Kerndocument. Van de voorstudie verschijnt een afzonderlijk rapport (Hoogenveen, Verkleij e.a., 1990).

Het onderhavige rapport vormt een terreinverkenning van met name de wiskundig-theoretische aspecten van concurrerende doodsoorzaken. Omdat onderzoek naar concurrerende doodsoorzaken gebaseerd is op de overlevingsduuranalyse (Engels: survival analysis) is eveneens van het laatstgenoemde onderwerp een terreinverkenning uitgevoerd. In het rapport wordt een aantal aspecten van beide onderwerpen, die regelmatig in de literatuur terugkeren, beschreven. Vooralnog zijn slechts voor een aantal van deze aspecten berekeningen uitgevoerd voor de specifieke Nederlandse situatie.

Mogelijke toepassingen van de theorie van concurrerende doodsoorzaken kunnen gevonden worden in het kader van het genoemde methodologie-project ten behoeve van de STG en de voorstudie ten behoeve van het Ontwerp-Kerndocument Gezondheidsbeleid (zie hoofdstuk 10). Aandacht kan worden geschonken aan de interactie van de doodsoorzaken hart- en vaatziekten, kanker, (verkeers)ongevallen, chronische ziekten (diabetes mellitus, CABA en reumatoïde arthritis) en AIDS.

INHOUDSOPGAVE

	blz.
Verzendlijst	ii
Voorwoord	iii
Inhoudsopgave	v
Summary	viii
Samenvatting	viii
Inleiding	1
1 Overlevingsfunctie en mortaliteitsdruk.	
1.1 Inleiding	3
1.2 De overlevingscurve en de daarin beschreven grootheden	4
1.3 De kansverdelingsfunctie van het overlijdenstijdstip	5
1.4 Overlevingsfunctie en mortaliteitsdruk	6
1.5 Voorbeelden van overlevingsfunctie en mortaliteitsdruk	7
1.6 Verschillende de sterfte beschrijvende grootheden	13
1.7 Vormen van sterftetabellen	14
1.8 Interpretatie in termen van frequentieverdeling	16
1.9 Een sterftetabel voor Nederland	16
1.10 Geparametriseerde mortaliteitsdrukken voor Nederland	18
2 De algemene aannemelijkheidsfunctie	
2.1 Inleiding	19
2.2 De gereduceerde steekproef methode en de Kaplan-Meier schatter	19
2.3 De actuariële methode	20
2.4 De aannemelijkheidsfunctie met gebruik van de mortaliteitsdruk	22
2.5 Type 1 censurering	24
2.6 Type 2 censurering	26
3 De partiële aannemelijkheidsfunctie	
3.1 Opsplitsing van de volledige aannemelijkheidsfunctie	27
3.2 De geneste structuur van de gebeurtenissen	28
3.3 De aannemelijkheidsfunctie geherformuleerd	29
3.4 De eerste term van de aannemelijkheidsfunctie	29
3.5 Eigenschappen van de partiële aannemelijkheidsfunctie	30
3.6 Tellingen en martingaal-theorie	31

4	Heterogeniteit	
4.1	Inleiding	35
4.2	Heterogeniteit en selectie	35
4.3	Formele introductie	36
4.4	Een voorbeeld van homogeniteit of heterogeniteit	38
4.5	Multiplicatieve heterogeniteit	39
4.6	Voorbeelden van multiplicatieve factoren	40
4.7	Heterogeniteit en partiële aannemelijkheidsfunctie	41
4.8	Afsluiting	42
5	Risico	
5.1	Inleiding	43
5.2	De accelerated-life beschrijving	43
5.3	De multiplicatief-risico beschrijving	44
5.4	De additief-risico beschrijving	45
5.5	Gegeneraliseerde lineaire modellen van risico	46
6	Concurrerende doodsoorzaken	
6.1	Inleiding	49
6.2	Beschrijving van het sterfteproces van een individu	49
6.3	De algemene aannemelijkheidsfunctie	50
6.4	De aannemelijkheidsfunctie onder type 1 censurering	51
6.5	De aannemelijkheidsfunctie onder type 2 censurering	52
6.6	De partiële aannemelijkheidsfunctie	52
6.7	Subject-intervallen	53
6.8	Een voorbeeld: het proportional hazards model	54
7	Virtuele overlijdenstijdstippen	
7.1	Inleiding	57
7.2	Introductie van het concept	57
7.3	Uitwerking van het concept	59
7.4	Sterke onderlinge onafhankelijkheid	60
7.5	Zwakke onderlinge onafhankelijkheid	60
7.6	Niet-identificeerbaarheid	61
7.7	Een voorbeeld van een keuzemogelijkheid	62
7.8	De algemene aannemelijkheidsfunctie	63
7.9	Vervangende en concurrerende doodsoorzaken	64
8	Het actuariële model	
8.1	Inleiding	65
8.2	Een overzicht van de gehanteerde grootheden	65
8.3	Beschrijving van de verschillende overlevings- en sterftekansen	66
8.4	Het schatten van de kansen uit de sterfte-aantallen	68
8.5	De kansverdelingen van de optredende sterfte	68
8.6	De schatters van de kansen	69
8.7	Uitwerking van de proportional hazards-aanname	71
8.8	Toepassing voor Nederland	72
8.9	Afsluiting	78

9	Geregistreerde sterftecijfers en overlevingsfuncties	
9.1	Inleiding	79
9.2	De voorwaardelijke overlevingsfractie als functie van de sterftegraad voor een stabiele bevolking	80
9.3	De voorwaardelijke overlevingskans als functie van de sterftegraad voor een groeiende bevolking	82
10	Conclusies en aanbevelingen	
10.1	Inleiding	85
10.2	Enkele conclusies op basis van de studie	85
10.3	Conclusies en aanbevelingen met betrekking tot vervolgonderzoeken	87
	Literatuurlijst	91
	Bijlage	
A	Berekening van de sterftetabel	
A.1	Inleiding	95
A.2	Het onder één noemer brengen van de gegevens	95
A.3	De berekening van de sterftetabel	97
A.4	Totale sterfte en verloren levensjaren	98
A.5	De sterfte na eliminatie van een doodsoorzaak	99
	Begrippenlijst	101

SUMMARY

The authors report on a literature search on survival analysis and competing risks. The subjects investigated are: (from survival analysis) survival functions, hazard functions and life tables, (total and partial) likelihood functions of the survival times, heterogeneity and risk functions to explain differences in survival times, (from competing risks) concurrent risks described by the two-dimensional stochastic variable survival time and cause of death, described by several latent stochastic survival times, and the actuarial model. Some of the subjects are illustrated with Dutch data.

SAMENVATTING

Dit rapport vormt het verslag van een literatuurstudie op het gebied van overlevingsduuranalyse en concurrerende doodsoorzaken. In dit rapport komen achtereenvolgens aan de orde: (m.b.t. de overlevingsduuranalyse:) overlevingsfuncties, mortaliteitsdruk en sterftetabellen en de dwarsdoorsnede- en cohort-interpretatie daarvan, (totale en partiële) aannemelijkheidsfuncties van het optreden van sterfte, heterogeniteit en risico om de verschillen tussen de sterfteprocessen van subgroepen te beschrijven, (m.b.t. concurrerende doodsoorzaken) concurrerende doodsoorzaken beschreven in termen van de stochastische variabelen overlijdenstijdstip en doodsoorzaak, in termen van virtuele overlijdenstijdstippen, en het actuariële model. Een aantal onderwerpen wordt uitgewerkt aan de hand van de Nederlandse situatie.

INLEIDING.

In dit rapport wordt verslag gedaan van een verkennend literatuuronderzoek op het terrein van concurrerende doodsoorzaken. Dit onderzoek was bedoeld om inzicht te krijgen in de theoretische achtergronden van dit concept. De uitwerking en toepassing van het concept zal mogelijk plaatsvinden in de vorm van vervolgonderzoeken.

Onderscheid wordt gemaakt tussen de concepten vervangende en concurrerende doodsoorzaken. In geval van vervangende doodsoorzaken wordt een vorm van onderlinge onafhankelijkheid van de doodsoorzaken verondersteld in termen van de doodsoorzaak-specifieke mortaliteitsdrukken; in geval van concurrerende doodsoorzaken een vorm van onderlinge afhankelijkheid. Onderlinge afhankelijkheid kan gemodelleerd worden door de doodsoorzaak-specifieke mortaliteitsdrukken te schrijven als functie van één of meer gemeenschappelijke risicofactoren.

Het onderwerp concurrerende doodsoorzaken vindt zijn basis in de overlevingsduuranalyse. Het verschil tussen beide vormt de specificering van de doodsoorzaak: bij concurrerende doodsoorzaken is de doodsoorzaak wél, bij overlevingsduuranalyse niet van belang. Veel onderwerpen uit de overlevingsduuranalyse zijn terug te vinden in beschrijvingen van concurrerende doodsoorzaken, bijvoorbeeld: overlevingsfunctie, mortaliteitsdruk, censurering, risico, en proportional hazards. Dit is de reden, dat in dit rapport ruim aandacht wordt besteed aan de overlevingsduuranalyse.

Overlevingsduuranalyse en onderzoek naar concurrerende doodsoorzaken zijn met name statistisch van aard. Het optreden van sterfte onder een cohort kan beschouwd worden als een stochastisch proces met een bijbehorende (onbekende) kansverdelingsfunctie. De vorm en eventuele parameter-waarden van deze kansverdelingsfuncties kunnen op grond van empirische sterftecijfers geschat worden met behulp van statistische technieken. Deze technieken behoren tot deelterreinen van de statistiek als actuariële wiskunde, risico analyse, en schattingstheorie.

In dit verslag komen de bovengenoemde onderwerpen aan de orde. In hoofdstuk 1 wordt een beschrijving gegeven van overlevingsfuncties. De overlevingsfunctie kan op twee manieren geïnterpreteerd worden: als de overlevingscurve van een cohort en als complement van de kansverdelingsfunctie van het overlijdenstijdstip van een individu. Tevens wordt een aantal voorbeelden van (parametrische) overlevingsfuncties gegeven. In hoofdstuk 2 worden twee algemene aannemelijkheidsfuncties afgeleid: die van de continu verdeelde overlijdenstijdstippen van de individuen van een cohort, en die van de discrete overlijdensaantallen voor gegeven tijdsintervallen. In hoofdstuk 3 wordt aandacht geschonken aan de herformulering van de algemene

aansamelijkheidsfunctie van de overlijdenstijdstippen als een partiële aansamelijkheidsfunctie.

In de hoofdstukken 4 en 5 worden de concepten heterogeniteit en risico geïntroduceerd. Heterogeniteit van een cohort uit zich in onderscheid in risico tussen de individuen. Het verschil tussen heterogeniteit en risico betreft de observeerbaarheid ervan: bij een heterogeniteit-benadering wordt het verschil in risico impliciet uit de overlijdenstijdstippen bepaald, bij een risico-benadering wordt uitgegaan van expliciet te meten specifieke risicovariabelen. In hoofdstuk 4 wordt de heterogeniteit van een cohort beschreven in termen van een stochastische toekemming van risico en een multiplicatieve risicofactor. In hoofdstuk 5 wordt een aantal modelmatige beschrijvingen van risico geïntroduceerd.

In de hoofdstukken 6 en 7 wordt aandacht besteed aan het verschijnsel concurrerende doodsoorzaken. In hoofdstuk 6 wordt de overleving van een individu beschreven in termen van twee stochastische variabelen: overlijdenstijdstip en doodsoorzaak. In hoofdstuk 7 wordt deze overleving beschreven in termen van virtuele overlijdenstijdstippen. Voor iedere specifieke doodsoorzaak kan een virtueel overlijdenstijdstip onderscheiden worden. Het minimum van deze overlijdenstijdstippen bepaalt het 'werkelijke' overlijdenstijdstip en 'de' doodsoorzaak. Hierbij wordt ruim aandacht besteed aan de (aanname van) (on-)afhankelijkheid van doodsoorzaken.

In hoofdstuk 8 wordt het actuariële model gepresenteerd, dat een toepassing vormt van de proportional hazards aanname met betrekking tot de verschillende (vervangende) doodsoorzaken. Een deel van de beschreven theorie wordt toegelicht aan de hand van de CBS-mortaliteitscijfers over 1986 en 1987 van Nederland. In hoofdstuk 9 tenslotte wordt beschreven hoe uit bestaande registraties onder verschillende aannames (de vorm en waarden van de parameters van) overlevingsfuncties geschat kunnen worden.

In de bijlage A wordt een aantal uitgevoerde berekeningen beschreven. Het gaat hierbij om de constructie van een sterftetabel voor Nederland en de constructie van alternatieve sterftetabellen na eliminatie van (elk van) de drie belangrijkste doodsoorzaken op basis van het actuariële model.

Deze inleiding wordt afgesloten met een notatie afspraak: stochastische variabelen worden vet gedrukt, vectoren onderstreept, en schatters voorzien van een $\hat{}$.

HOOFDSTUK 1 Overlevingsfunctie en mortaliteitsdruk.

Paragraaf 1.1 Inleiding.

Sterftetabellen (Engels: life tables) beschrijven de leeftijds-afhankelijke sterftesantallen en daarmee ook de aantallen nog levenden. In overlevingsduuranalyse worden de sterftepatronen van groepen individuen (c.q. levende organismen) onderzocht. Bij sterftetabellen en overlevingsduuranalyse staat het overlijdenstijdstip van individuen centraal. Dit overlijdenstijdstip kan beschouwd worden als een stochastische grootheid. Het in waarschijnlijkheid optreden van overlijden wordt beschreven door de bijbehorende kansverdelingsfunctie. Het complement van deze functie wordt de overlevingsfunctie (Engels: survival function) genoemd. Deze functie kan ook geïnterpreteerd worden als de overlevingscurve van een geboorte-cohort: het beschrijft voor iedere leeftijd het aantal nog levenden als fractie van het beginaantal bij de geboorte.

Een aan de overlevingsfunctie gelijkwaardige grootheid is de mortaliteitsdruk (Engels: hazard function, hazard rate). Deze beschrijft voor iedere leeftijd de relatieve verandering van de overlevingsfunctie. Omgekeerd beschrijft de overlevingsfunctie (na transformatie) de cumulatieve mortaliteitsdruk.

In dit hoofdstuk wordt de overlevingsfunctie langs de twee genoemde wegen geïntroduceerd: op basis van de overlevingscurve van een geboorte-cohort en als complement van de kansverdelingsfunctie van het overlijdenstijdstip. Hierbij wordt een aantal van de overlevingsfractie en mortaliteitsdruk af te leiden grootheden genoemd. Vervolgens wordt een aantal voorbeelden van overlevingsfuncties gegeven, die volgen uit specifieke aannames met betrekking tot de vorm van de overlevingsfractie of de mortaliteitsdruk. Deze afgeleide grootheden spelen o.a. een rol bij de koppeling van overlevingsfuncties aan bestaande registraties in de vorm van sterftetabellen.

Strikt genomen is de dubbele interpretatie van de overlevingsfunctie (een geschaalde overlevingscurve van een geboorte-cohort resp. het complement van de kansverdelingsfunctie van het overlijdenstijdstip) niet juist. In het eerste geval beschrijft de overlevingsfunctie het gerealiseerd sterftepatroon van een geboorte-cohort, in het tweede geval de te toetsen vorm van de kansverdelingsfunctie van een onbekende stochastische grootheid, namelijk het overlijdenstijdstip. Deze twee functies zijn uiteraard niet identiek. In het algemeen kan gesteld worden, dat de genoemde dubbele interpretatie van de overlevingsfunctie van twee vooronderstellingen uitgaat: geen censurering en een homogene onderzoeksgroep. Met censurering wordt bedoeld, dat een individu om een onbekende reden uit een onderzoeksgroep onttrokken wordt (zie paragraaf 2.5 e.v.). Onder een homogene

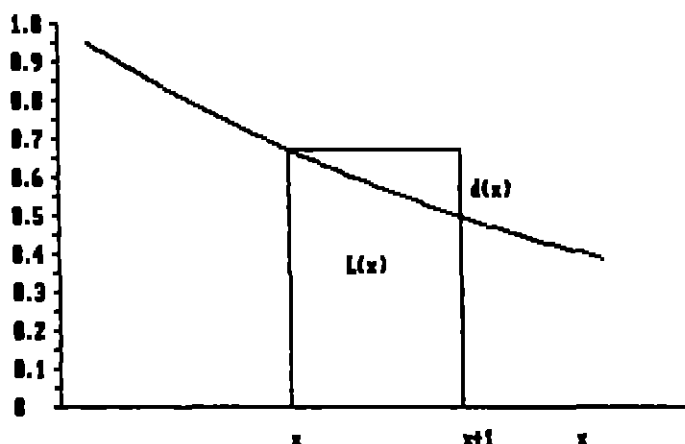
onderzoeksgroep wordt in dit verband een verzameling onderling onafhankelijke identieke individuen verstaan (zie hoofdstuk 4).

Paragraaf 1.2 De overlevingscurve en de daarin beschreven grootheden.

De overlevingsfunctie kan geïntroduceerd worden vanuit de overlevingscurve van een geboorte-cohort. Een groep van identieke personen wordt vanaf de geboorte gevolgd en voor iedere leeftijd x wordt het aantal nog levende individuen $l(x)$ genoteerd. Uittrede uit het cohort is slechts mogelijk in de vorm van sterfte (geen censurering, zie paragraaf 2.5). Het aantal $l(x)$ wordt zo geschaald, dat het beginaantal gelijk aan 1 is. De zo geconstrueerde functie beschrijft voor iedere leeftijd x de overlevingsfractie $l(x)$ van het cohort. Op leeftijd $x=0$ is deze fractie 1 en neemt vervolgens af (monotoon niet-stijgend) met de leeftijd.

Een voorbeeld van een dergelijke overlevingsfunctie is gegeven in figuur 1. Hierin is een overlevingsfunctie $l(x)$ uitgezet tegen de leeftijd x . In het volgende schema worden deze grootheden genoemd en zo nodig formeel gedefinieerd als functie van de overlevingsfunctie.

symbool	omschrijving
x	leeftijd c.q. leeftijdsinterval $[x, x+1)$
$[x, x+1)$	leeftijdsinterval
$l(x), S(x)$	overlevingsfractie op leeftijd x
$L(x)$	gemiddelde overlevingsfractie tussen x en $x+1$, $= \int_0^1 l(x+s) ds$
$d(x)$	afname van overlevingsfractie, $= l(x) - l(x+1)$



Figuur 1: Een overlevingsfunctie.

Noot: x = leeftijd, $l(x)$ = overlevingsfunctie

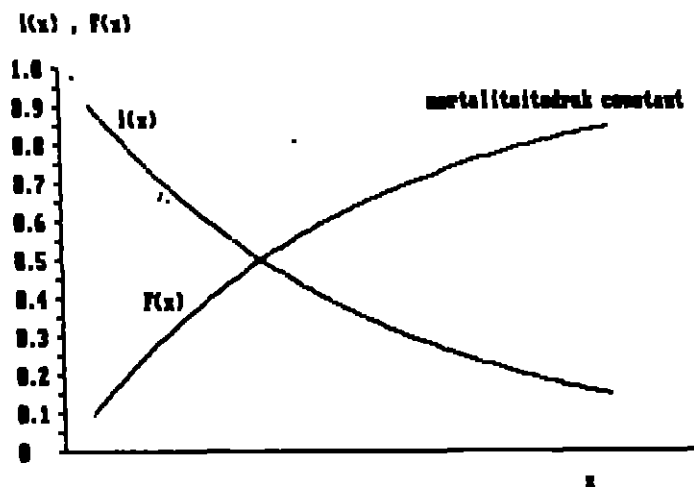
Paragraaf 1.3 De kansverdelingsfunctie van het overlijdenstijdstip.

De overlevingsfunctie kan eveneens geïntroduceerd worden vanuit de kansverdelingsfunctie van het stochastische overlijdenstijdstip x van een willekeurig individu. Deze functie geeft voor iedere leeftijd de kans, dat het individu voor dit tijdstip overlijdt:

$$F(x) = \Pr(x \leq x)$$

In figuur 2 wordt de kansverdelingsfunctie van een continu verdeeld overlijdenstijdstip en de daaraan complementaire overlevingsfunctie gegeven. Deze functies en de daaruit af te leiden grootheden worden analoog aan paragraaf 1.2 verbaal genoemd en formeel gedefinieerd.

symbool	omschrijving
x	leeftijd
x	het stochastische overlijdenstijdstip
$F(x)$	kansverdelingsfunctie van x , $= \Pr(x \leq x)$
$f(x)$	kansdichtheidsfunctie van x , $= F'(x)$
$l(x)$	de overlevingsfunctie, $= \Pr(x > x) = 1 - F(x)$



Figuur 2: De kansverdelingsfunctie van een overlijdenstijdstip.

Noot: x = leeftijd, $l(x)$ = overlevingsfunctie,
 $F(x)$ = kansverdelingsfunctie

Het overlijdenstijdstip vormt een stochastische variabele x , waarvan de functie $F(x)$ de kansverdelingsfunctie vormt. Deze variabele kan via de overlevingsfunctie in verband gebracht worden met het sterftepatroon van een geboorte-cohort. Op grond van deze interpretatie kan een aantal andere uit de overlevingsfunctie af te leiden grootheden beschreven worden:

$e(x)$	de op leeftijd x resterende levensverwachting
$h(x)$	de mortaliteitsdruk
$H(x)$	de cumulatieve mortaliteitsdruk, ofwel: het aanwezige risico
$p(x), q(x)$	de voorwaardelijke eenjaars overlevings- resp. sterftefractie

$$e(x) = \int_x^{\infty} (s-x) dF(s|x) ds = \int_x^{\infty} l(s) ds / l(x)$$

$$h(x) = \lim_{h \rightarrow 0} \frac{\Pr(x < x+h | x \geq x)}{h} = f(x) / l(x) = - \frac{d}{dx} \ln l(x)$$

$$H(x) = \int_0^x h(s) ds = - \ln l(x)$$

$$q(x) = \Pr(x < x+1 | x \geq x) = d(x) / l(x)$$

$$p(x) = \Pr(x \geq x+1 | x \geq x) = l(x+1) / l(x) = 1 - q(x)$$

Paragraaf 1.4 . Overlevingsfunctie en mortaliteitsdruk.

In paragraaf 1.3 zijn verschillende grootheden gedefinieerd in termen van de overlevingsfunctie. Daarnaast is het ook mogelijk om als basisgrootte de mortaliteitsdruk (voor continue overlevingscurven) of de voorwaardelijke eenjaars-overlevingsfractie (voor discrete overlevingscurven) te nemen. De definitie van de overlevingsfunctie in termen van de mortaliteitsdruk c.q. de voorwaardelijke eenjaars-overlevingsfracties luidt:

$$l(x) = \exp - \int_0^x h(s) ds = \exp - H(x) \text{ resp.}$$

$$l(x) = \prod_{i=0, \dots, x-1} p(i)$$

Analoog kan de voorwaardelijke eenjaars-overlevingsfractie geschreven worden als een functie van de continue mortaliteitsdruk:

$$p(x) = \exp - \int_x^{x+1} h(s) ds = \exp H(x) - H(x+1)$$

Paragraaf 1.5 Voorbeelden van overlevingsfunctie en mortaliteitsdruk.

De vorm van de (continue) overlevingsfunctie c.q. de mortaliteitsdruk kan vastgelegd worden met behulp van een geparametriseerd functievoorschrift. In deze paragraaf wordt een aantal voorbeelden van geparametriseerde overlevingsfuncties gegeven (zie bijvoorbeeld ook Cox ea, 1984), zowel in formulevorm als (voor een specifieke parameterkeuze) in grafische vorm. Twee groepen van voorbeelden komen aan bod.

Voor de eerste groep wordt voor alle leeftijdsjaren x de vorm van de mortaliteitsdruk $h(x)$ gegeven of een daarmee gelijkwaardige grootheid. Hieruit kan voor ieder leeftijdsjaar x de overlevingsfractie $l(x)$ bepaald worden. Voor de tweede groep worden gegeven: de overlevingsfracties $l(x)$ resp. $l(x+1)$ aan het begin x en het eind $x+1$ van een leeftijdsinterval $[x, x+1)$ en de vorm van de (continue) sterftefractie $q(x+dx)$ of mortaliteitsdruk $h(x+dx)$ over het interval $(0 < dx < 1)$. Hieruit kunnen de overlevingsfracties voor alle leeftijden binnen het interval bepaald worden.

Voor elk van de voorbeelden wordt de vorm (formule) van een aantal grootheden beschreven, die de overlevingsfunctie karakteriseren. De meeste van deze grootheden zijn reeds genoemd in de paragrafen 1.2 en 1.3. Een nieuwe, meer algemeen op kansverdelingsfuncties geöriënteerde grootheid is:

$$g(s) \quad \text{de karakteristieke functie van de kansverdeling } F,$$

$$= E(\exp^{-sx}) = \int_0^{\infty} e^{-sx} dF(x)$$

De eerste groep van voorbeelden.

Voorbeeld 1: de exponentiële verdeling (figuur 3).

$$h(x) = c \quad \text{constante mortaliteitsdruk}$$

Daaruit volgt:

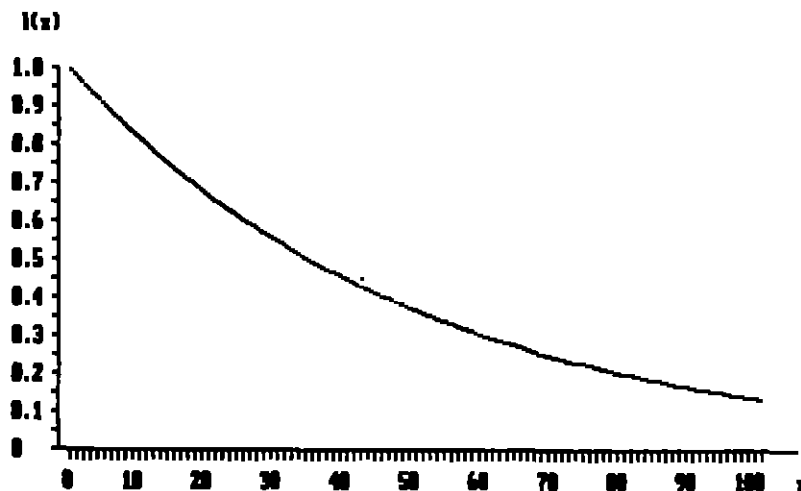
$$H(x) = c x \quad \text{lineair toenemend risico}$$

$$l(x) = \exp(-cx) \quad \text{exponentieel afnemende fractie}$$

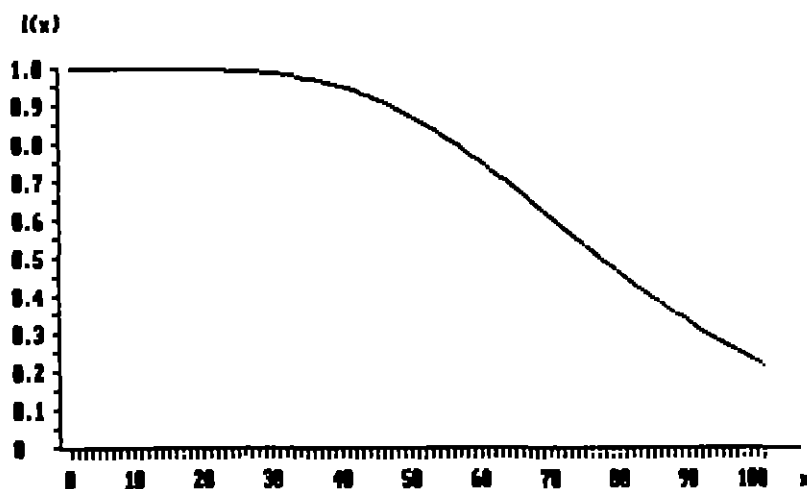
$$e(x) = 1 / c \quad \text{constante resterende levensverwachting}$$

$$g(t) = \frac{c}{c + s} \quad \text{karakteristieke functie}$$

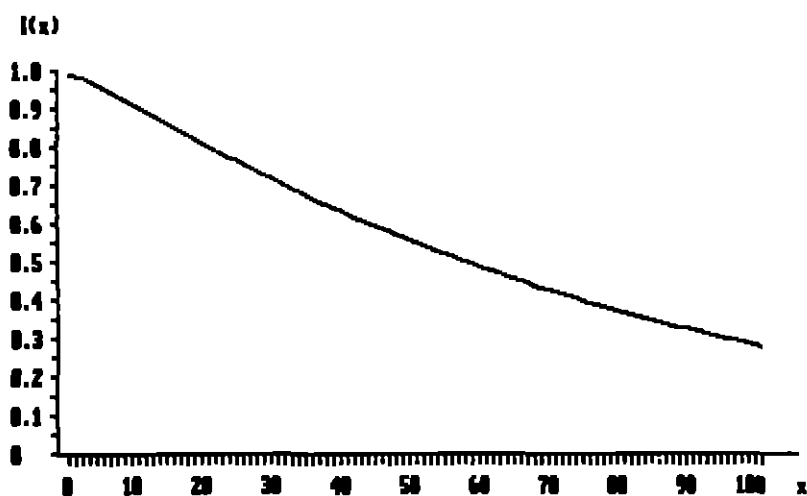
Een constante mortaliteitsdruk houdt in, dat geen veroudering optreedt en dat de (resterende) levensverwachting voor elke leeftijd gelijk is.



Figuur 3: De exponentiële verdeling.



Figuur 4: De gamma-verdeling.



Figuur 5: De Weibull-verdeling.

Noot: x = leeftijd, $l(x)$ = overlevingsfunctie

Voorbeeld 2: de gamma- (Erlang-) verdeling (figuur 4).

$$f(x) = c * cx^{n-1} \exp(-cx) / (n-1)!$$

Daaruit volgt:

$$\text{gem}(x) = n / c \quad \text{levensverwachting bij geboorte}$$

$$\text{modus}(x) = n-1 / c \quad \text{modus neigt naar gemiddelde}$$

$$g(t) = \left(\frac{c}{c + t} \right)^n \quad \text{karakteristieke functie}$$

Een gamma-verdeeld overlijdenstijdstip kan beschouwd worden als de som van een reeks exponentieel verdeelde tijdstippen. Voorafgaand aan de sterfte worden verschillende stadia met exponentieel verdeelde duur (verblijfstijd, Engels: sojourn time) doorlopen.

Voorbeeld 3: de Weibull-verdeling (figuur 5).

$$h(x) = a * b * x^{b-1} \quad a, b > 0$$

met:

- $b < 1$: dalende mortaliteitsdruk
- $b = 1$: constante mortaliteitsdruk
- $b > 1$: stijgende mortaliteitsdruk,

Daaruit volgt:

$$l(x) = \exp(-a * x^b) \quad \text{overlevingsfractie Weibull-verdeeld}$$

Voorbeeld 4: de Gompertz-verdeling (figuur 6).

$$h(x) = a * b * \exp(bx) \quad a, b > 0 \quad \text{exponentieel toenemende druk}$$

Daaruit volgt:

$$l(x) = \exp(-a * (\exp(bx) - 1))$$

Voorbeeld 5: de Makham-verdeling (figuur 7).

$$h(x) = a + b * \exp(dx) \quad a, b, d > 0$$

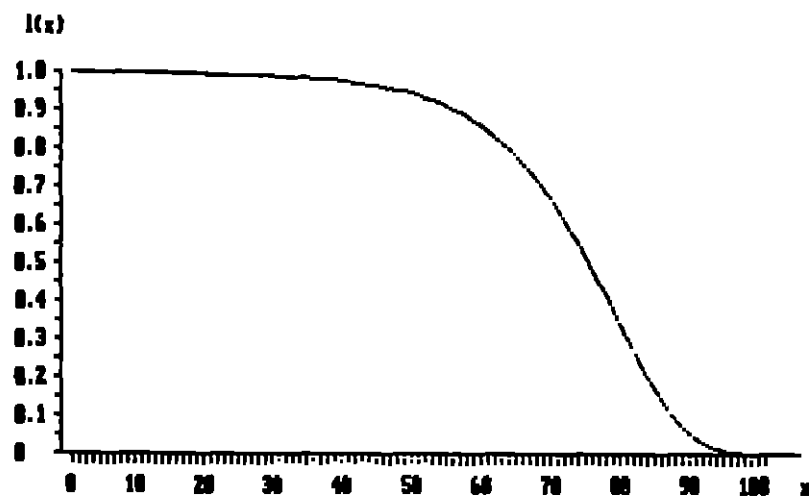
de mortaliteitsdruk neemt vanaf een drempelwaarde exponentieel toe

Daaruit volgt:

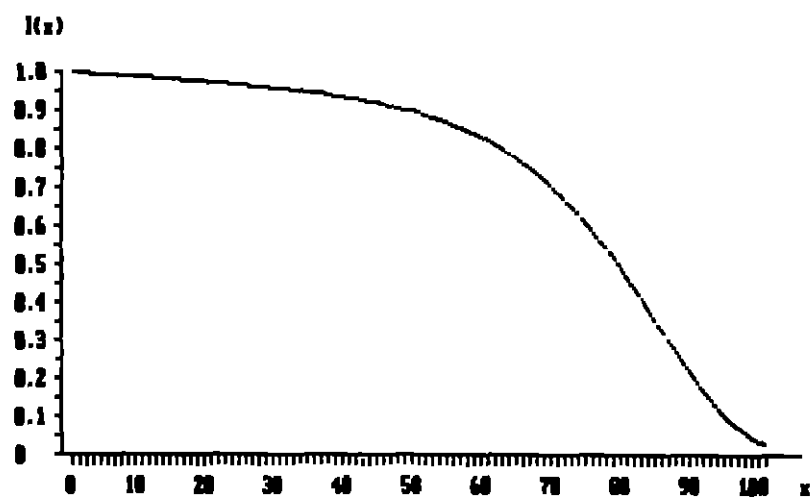
$$l(x) = s * g^{(c^x - 1)}$$

met:

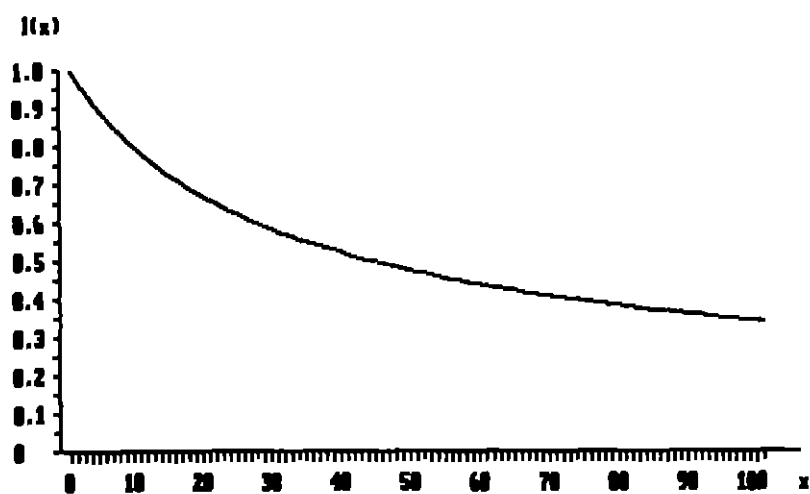
$$s = \exp(-a) < 1, \quad g = \exp(-b/c) < 1, \quad c = \exp(d) > 1$$



Figuur 6: De Gompertz-verdeling.



Figuur 7: De Makeham-verdeling.



Figuur 8: De Pareto-verdeling.

Noot: x = leeftijd, $l(x)$ = overlevingsfunctie

Voorbeeld 6: de samengestelde exponentiële verdeling.

$$h(x|c) = c \quad \text{conditionele constante mortaliteitsdruk}$$

Daaruit volgt:

$$f(x) = \int f(x|c) g(c) dc \quad \text{onvoorwaardelijke kansdichtheid}$$

Zij de mortaliteitsdruk (bijvoorbeeld) gamma-verdeeld met parameters C en n , dan volgt daaruit:

$$f(x) = n C^n / (x+C)^{n+1} \quad \text{Pareto verdeling (figuur 8)}$$

$$l(x) = C^n / (x+C)^n$$

$$h(x) = n / x+C$$

De tweede groep van voorbeelden.

Als grenzen voor het leeftijdsinterval worden de leeftijden x resp $x+1$ gekozen met bijbehorende gegeven overlevingsfracties $l(x)$ resp. $l(x+1)$. De aannames betreffen de vorm van de (continue) sterftefractie c.q. de mortaliteitsdruk over het interval. Deze voorbeelden geven aan, hoe op basis van een discrete overlevingsfunctie een continue, stuksgewijs continue differentieerbare overlevingsfunctie geconstrueerd kan worden. Voor alle voorbeelden geldt voor de tijdsparameter: $0 \leftarrow dx \leftarrow 1$.

Voorbeeld 7: uniforme verdeling (figuur 9).

$$q(x, x+dx) = dx * q(x)$$

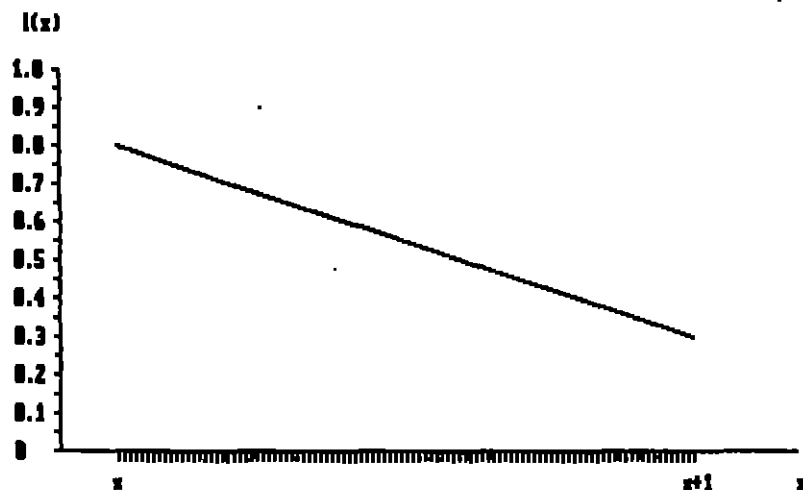
Daaruit volgt:

$$l(x+h) = (1-dx) l(x) + dx l(x+1) \quad \text{lineaire overlevingsfractie}$$

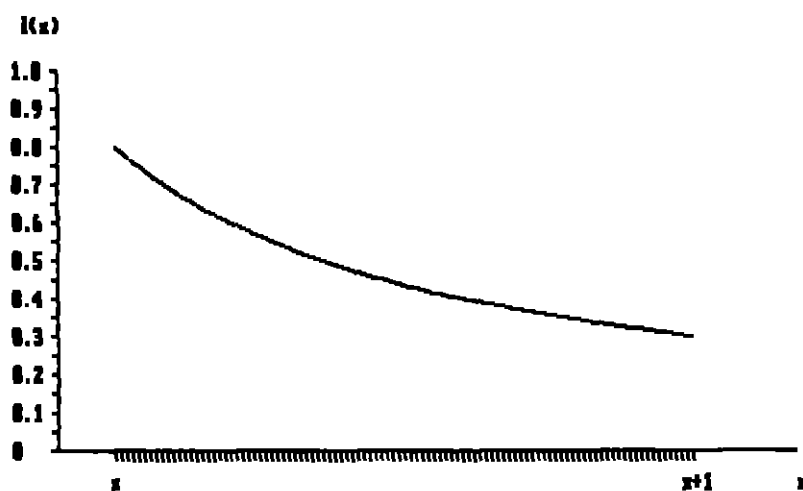
$$h(x+dx) = q(x) / 1 - dx*q(x) \quad \text{toenemende mortaliteitsdruk}$$

$$q(x+dx1, x+dx2) = (dx2-dx1) * d(x) / l(x+dx1)$$

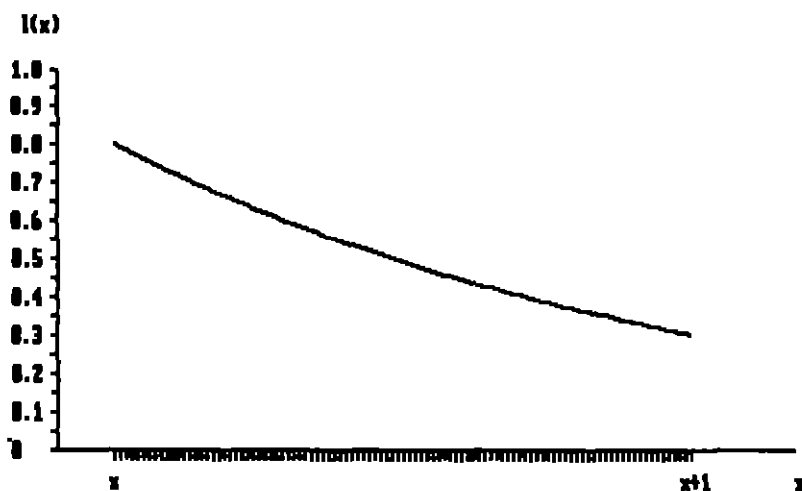
resterende sterfte uniform verdeeld



Figuur 9: De uniforme verdeling.



Figuur 10: De Baldacci-verdeling.



Figuur 11: De exponentiële verdeling.

Not: x = leeftijd, $l(x)$ = overlevingsfunctie

Voorbeeld 8: Balducci-verdeling (figuur 10).

$$q(x+dx, x+1) = (1-dx) * q(x)$$

Daaruit volgt:

$$l(x+dx) = l(x) * l(x+1) / (dx * l(x) + (1-dx) * l(x+1))$$

convexe overlevingsfractie

$$h(x+dx) = q(x) / 1 - (1-dx) * q(x) \quad \text{afnemende mortaliteitsdruk}$$

$$q(x, x+dx) = dx * q(x) / (1 - (1-dx) * q(x))$$

Voorbeeld 9: exponentiële verdeling (figuur 11).

$$h(x+dx) = c = \ln \frac{l(x)}{l(x+1)} \quad \text{constante mortaliteitsdruk}$$

Daaruit volgt (met $p(x, x+1) = \exp(-c)$):

$$l(x+dx) = l(x) e^{-cdx} = l(x) p(x)^{dx} \quad \text{convexe overlevingsfractie}$$

$$p(x, x+dx) = p(x)^{dx} \quad \text{convexe relatieve overlevingsfractie}$$

Paragraaf 1.6 Verschillende de sterfte beschrijvende grootheden.

Verschillende grootheden kunnen onderscheiden worden, die een beschrijving geven van de voorwaardelijke relatieve sterfte, d.w.z. van de absolute sterfte, gecorrigeerd voor de bevolkingsgrootte.

Deze grootheden zijn:

$$q(x) = d(x) / l(x) \quad \text{de sterftefractie}$$

$$m(x) = d(x) / L(x) \quad \text{de centrale sterftegraad}$$

$$h(x) = - l'(x) / l(x) \quad \text{de mortaliteitsdruk}$$

De grootheid $d(x)$ geeft de absolute sterfte gedurende het leeftijds-interval $[x, x+1)$ weer, de grootheden $q(x)$ en $m(x)$ de relatieve sterfte. Voor $q(x)$ wordt de absolute sterfte gedeeld door de overlevingsfractie op leeftijd x , voor $m(x)$ door die met leeftijd tussen x en $x+1$. Deze grootheid $m(x)$ vormt een tussenschakel tussen overlevingsfuncties en bestaande registraties van sterftegegevens (zie hoofdstuk 9).

De mortaliteitsdruk $h(x)$ komt overeen met de limiet van de sterftekans of sterftegraad, indien in plaats van het leeftijdsinterval $[x, x+1)$ het infinitesimaal kleine interval $[x, x+h)$ ($h \rightarrow 0$) genomen wordt.

Tussen de genoemde grootheden bestaan de volgende relaties:

$$q(x) = m(x) / (1 + (1-a(x)) \cdot m(x))$$

$$m(x) = q(x) / (1 - (1-a(x)) \cdot q(x))$$

met:

$$a(x) = (L(x) - l(x+1)) / d(x)$$

de resterende levensverwachting op leeftijd x van hen, die voor leeftijd $x+1$ sterven, en dus:

$$q(x) \leftarrow m(x) \leftarrow q(x) / 1 - q(x)$$

$$m(x) / 1 + m(x) \leftarrow q(x) \leftarrow m(x)$$

De laatstgenoemde ongelijkheden worden een gelijkheid, indien alle sterfte aan het begin van het interval resp. alle sterfte aan het einde van het interval plaatsvindt. Indien $l(x)$ lineair wordt verondersteld geldt $a(x) = 0.5$.

Paragraaf 1.7 Vormen van sterftetabellen.

Sterftetabellen bevatten (voor een reeks achtereenvolgende waarnemingstijdstippen) de aantallen levende personen per leeftijdsjaar (zie bijvoorbeeld Chiang, 1968). Daarnaast worden meestal ook andere, uit de aantallen levende personen af te leiden aantallen genoteerd. Een sterftetabel kan bijvoorbeeld de volgende kolommen bevatten:

kolom 1 het leeftijdsinterval $[x, x+1)$

kolom 2 de sterftefractie

kolom 3 het aantal personen met leeftijd tussen x en $x+1$ aan het begin van of halverwege het jaar

kolom 4 het aantal gedurende een waarnemingsperiode gestorven personen met leeftijd tussen x en $x+1$

kolom 5 de fractie van het leeftijdsinterval, dat gemiddeld geleefd is

kolom 6 de levensverwachting op leeftijd x

Het is mogelijk om aan de hand van verschillende criteria sterftetabellen in te delen.

Cohort- of dwarsdoorsnede-vorm

Een cohort is een vaste groep personen in de tijd gevolgd, zonder vorm van instroom via bijvoorbeeld geboorte of immigratie, maar wel met mogelijke uitstroom in de vorm van sterfte. De optredende sterfte wordt afhankelijk van het tijdstip van waarneming in een cohort-sterftetabel vastgelegd. In de praktijk wordt meestal een geboorte-cohort gekozen: een groep personen met (ongeveer) hetzelfde geboortjaar en dus, voor een vast waarnemingstijdstip, met (ongeveer) dezelfde leeftijd. De sterfte afhankelijk van het tijdstip van waarneming komt dan overeen met die afhankelijk van de leeftijd.

Daarnaast kan op een vast tijdstip een dwarsdoorsnede van de bevolking, verdeeld naar leeftijd, opgemaakt worden. Van deze dwarsdoorsnede wordt de leeftijdsafhankelijke sterfte gedurende (zeg) een jaar genoteerd in een tabel, de dwarsdoorsnede-vorm.

Continu of discreet

Een sterftetabel beschrijft de sterfte, verdeeld naar de leeftijd. Deze tabel heet continu of discreet, indien deze verdeling continu resp. discreet is.

Een discrete sterftetabel kan beschouwd worden als de discrete vertaling (discretisering) van een continue tabel: de sterfte, welke continu van de leeftijd afhangt, wordt toegewezen aan een eindig aantal intervallen van leeftijdsjaren. Omgekeerd kan op basis van een discrete sterftetabel een continue tabel geconstrueerd worden: de discrete tabel geeft per leeftijdsinterval de randvoorwaarden, waaraan de continue tabel moet voldoen.

Compleet of ingekort

Een discrete sterftetabel, waarin dus de leeftijd discreet verdeeld is, kan compleet of ingekort (Engels: abridged) zijn. In het eerste geval is de leeftijd in afzonderlijke jaren verdeeld, in het tweede in meerjaars-groepen.

Een bevolking(sopbouw) heet stationair (stabiel), indien de samenstelling ervan, verdeeld naar leeftijd (en geslacht), niet verandert in de loop van de waarnemingstijd. Voor een stationaire bevolking zijn de sterftetabellen in cohort- en dwarsdoorsnede-vorm identiek en onafhankelijk van het waarnemingstijdstip. Voor een niet-stationaire bevolking zijn de sterftetabellen verschillend en hangen deze wel af van het waarnemingstijdstip.

Paragraaf 1.8 Interpretatie in termen van frequentieverdeling.

De overlevingscurve beschrijft voor ieder leeftijdsjaar het aantal personen van een geboorte-cohort, dat nog in leven is. Deze overlevingscurve kan (via schaling) geïnterpreteerd worden als een overlevingsfunctie. In aansluiting hierop zijn de verschillende uit een dergelijke functie af te leiden grootheden beschreven in termen van de veranderingen over de tijd en daarmee leeftijd. Een dergelijke overlevingsfunctie sluit aan op de cohort-vorm sterftetabel.

Het is echter ook mogelijk om de (geschaalde) frequentieverdeling van een bevolking(sgroep) op een zeker ogenblik, een dwarsdoorsnede, te interpreteren als een overlevingsfunctie. Een dergelijke overlevingsfunctie beschrijft de momentane toestand van de bevolking en sluit aan op de sterftetabel in dwarsdoorsnede-vorm, zoals die meestal gehanteerd wordt bij registraties. Omdat de bevolkingsopbouw in het algemeen niet stabiel is, zal deze (geschaalde) frequentieverdeling niet overeenkomen met een overlevingsfunctie op basis van een geboorte-cohort. Daarom moeten methoden gevonden worden om sterftetabellen om te zetten van de dwarsdoorsnede-vorm in de cohort-vorm.

Een aantal grootheden van een overlevingsfunctie kan ook geïnterpreteerd worden in termen van een dwarsdoorsnede:

$l(x)$	deel van de bevolking met leeftijd x ($l(0)=1$)
$L(x)$	deel van de bevolking met leeftijd tussen x en $x+1$
$d(x)$	deel van de bevolking, dat met leeftijd tussen x en $x+1$ gedurende een jaar sterft

Paragraaf 1.9 Een sterftetabel voor Nederland.

In deze paragraaf wordt een voorbeeld gegeven van een dwarsdoorsnede-sterftetabel voor Nederland. Hierbij wordt uitgegaan van het gemiddelde van de sterftefracties over de jaren 1986 en 1987 (CBS, 1986, 1987). Deze fracties hebben betrekking op vijfjaars leeftijdsklassen. De jongste klasse, van 0 tot 5 jaar, is fijner verdeeld. In tabel 1 is een aantal grootheden van een sterftetabel uitgezet, namelijk: de gegeven sterftefractie q , en het aantal overlevenden l , beide met leeftijd x , het aantal overlevenden L , behorend tot de leeftijdsklasse, waarvan x het beginjaar is, het aantal mensjaren T nog te leven door de overlevenden met leeftijd x , en de resterende levensverwachting op leeftijd x . De verschillende uitgevoerde berekeningen zijn beschreven in bijlage A.

Tabel 1: Een sterftetabel voor Nederland.

x	q	l	L	T	e
mannen					
0	.00728	100.000	99.636	7.354.890	73,6
1	.00176	99.272	99.184	7.255.254	73,1
2	.00047	99.097	297.079	7.156.070	72,2
5	.00021	98.956	494.512	6.858.991	69,3
10	.00021	98.849	493.981	6.364.479	64,4
15	.00051	98.743	493.087	5.870.498	59,5
20	.00074	98.492	491.542	5.377.411	54,6
25	.00072	98.125	489.738	4.885.869	49,8
30	.00083	97.770	487.837	4.396.131	45,0
35	.00118	97.365	485.392	3.908.294	40,2
40	.00189	96.792	481.675	3.422.902	35,4
45	.00329	95.878	475.468	2.941.226	30,7
50	.00580	94.309	464.781	2.465.758	26,2
55	.01022	91.603	446.551	2.000.978	21,8
60	.01731	87.017	416.893	1.554.427	17,9
65	.02883	79.740	371.573	1.137.534	14,3
70	.04669	68.889	307.821	765.961	11,1
75	.07496	54.239	227.442	458.140	8,5
80	.11416	36.738	141.944	230.698	6,3
85	.20288	20.040	88.754	88.754	4,4
vrouwen					
0	.00565	100.000	99.717	8.017.303	80,2
1	.00147	99.434	99.361	7.917.585	79,6
2	.00037	99.288	297.696	7.818.224	78,7
5	.00015	99.176	495.689	7.520.528	75,8
10	.00018	99.099	495.274	7.024.839	70,9
15	.00023	99.010	494.760	6.529.566	66,0
20	.00032	98.894	494.068	6.034.805	61,0
25	.00038	98.773	493.192	5.540.737	56,1
30	.00057	98.543	492.009	5.047.546	51,2
35	.00078	98.260	490.345	4.555.536	46,4
40	.00124	97.878	487.876	4.065.191	41,5
45	.00207	97.272	483.850	3.577.315	36,8
50	.00323	96.267	477.469	3.093.465	32,1
55	.00517	94.720	467.543	2.615.996	27,6
60	.00781	92.297	452.609	2.148.453	23,3
65	.01225	88.746	430.466	1.695.884	19,1
70	.02069	83.440	396.496	1.265.378	15,2
75	.03730	75.158	343.263	868.882	11,6
80	.06819	62.147	264.510	525.619	8,5
85	.15430	43.657	261.109	261.109	6,0

Noot: x = beginjaar van leeftijdscategorie, q = sterftefractie, l = overlevingsfractie, L = levensjaren binnen categorie, T = resterende levensjaren, e = resterende levensverwachting. Bron: CBS, 1986, 1987

**Paragraaf 1.10 Geparmetriseerde mortaliteitsdrukken voor
Nederland.**

Van een aantal geparmetriseerde mortaliteitsdrukken (zie paragraaf 1.5) is de fit voor de Nederlandse situatie bepaald. Daarbij zijn de leeftijds- en geslachtsspecifieke sterftefracties q geïnterpreteerd als mortaliteitsdrukken. De parameterwaarden worden zo gekozen, dat de kwadraatsom van de verschillen tussen de functiewaarden en de vaargenomen mortaliteitsdrukken voor de verschillende leeftijds-groepen minimaal is. In tabel 2 worden de resultaten van de functionele benaderingen beschreven. De kwadratische afwijking tussen de werkelijke mortaliteitsdrukken en de benaderingen wordt gegeven als verhoudingsgetal ten opzichte van die voor een constante mortaliteitsdruk voor mannen.

Tabel 2: Geschatte parameters van de mortaliteitsdruk.

	mort.druk		parameterwaarden		kwadratische afwijking	
constant c	M		1.3 E-2		1	
	V		1.1 E-2		0.8	
Weibull $a*b*x^{b-1}$	M		2.0 E-3	1.047	0.2	
	V		2.2 E-3	1.040	0.3	
Gompertz $a*b*e^{bx}$	M		5.7 E-4	9.6 E-2	6 E-3	
	V		2.7 E-4	10.0 E-2	3 E -2	
Makeham $a+b*e^{dx}$	M		8.9 E-4	4.0 E-5	1.0 E-1	5 E-3
	V		2.1 E-4	4.0 E-7	1.5 E-1	5 E-3

Noot: a,b,c,d = parameterwaarden, x = leeftijd, M = mannen,
V = vrouwen, E = 10-macht notatie voor reële getallen

Hieruit blijkt, dat de mortaliteitsdruk volgend uit de Makeham-overlevingsfunctie de gegeven sterftefracties het beste benadert.

HOOFDSTUK 2 Schattingsmethoden.

Paragraaf 2.1 Inleiding.

Verschillende methoden kunnen onderscheiden worden om een (parameter vrije) overlevingsfunctie c.q. de parameterwaarden van een gegeven (geparametriseerde) overlevingsfunctie te schatten. Tot de eerste groep van (parameter vrije) schattingsmethoden kunnen gerekend worden: de gereduceerde steekproef methode, de actuariële methode, en de methode van Kaplan-Meier. Bij de tweede groep van (parametrische) methoden wordt de meest aannameijke schatter(s) van de onbekende parameter(s) van de gegeven overlevingsfunctie bepaald door de aannameijkheidsfunctie van de (gerealiseerde) overlijdens-tijdstippen c.q. sterfteaantallen te maximaliseren. Beide groepen van methoden worden achtereenvolgens behandeld. Het hoofdstuk eindigt met een beschrijving van de gevolgen van censurering voor de beschreven schattingsfuncties (zie bijvoorbeeld Chiang, 1968, en Moeschberger ea, 1971).

Paragraaf 2.2 De gereduceerde steekproef methode en de Kaplan-Meier schatter.

Bij de gereduceerde steekproef methode wordt de overlevingscurve van een (homogeen verondersteld) (geboorte-)cohort gebruikt als een schatting van de overlevingsfunctie van elk van de individuen van het cohort. Gebruik wordt gemaakt van de volgende grootheden:

- $I(x)$ indicatorfunctie voor het nog in leven zijn
- = 0 ($x < 0$)
- = 1 ($x \geq 0$)
- $N = I(0)$ de verzameling individuen van een cohort
- i index met betrekking tot de individuen
- t_i het gerealiseerde sterftetijdstip van individu i

Zonder beperking van de algemeenheid wordt verondersteld, dat de individuen naar toenemend tijdstip van overlijden geordend zijn. De gereduceerde steekproef schatter luidt:

$$\hat{S}(x) = \sum_{i \in N} I(t_i - x) / \#(N)$$

Deze schatter vormt een stapfunctie met, indien alle overlijdens-tijdstippen verschillend zijn (Engels: no ties), telkens stapgrootte $1/\#(N)$. De statistische eigenschappen van deze schatter zijn:

$$E \hat{S}(x) = \sum (I) E I(t_i - x) / \#(N) = S(x)$$

$$\text{var } \hat{S}(x) = \sum (I/N) \text{var } I(t_i - x) / \#(N)^2 = S(x) * (1 - S(x)) / N$$

De gereduceerde steekproefschatter kan worden herschreven tot een produktreeks:

$$\hat{S}(t) = \prod_{t_j \leq t} \frac{N_j}{N_{j-1}}$$

met:

$$N_j = \sum (k) I(t_k - t_j)$$

Deze schatter wordt de Kaplan-Meier schatter genoemd. De afzonderlijke termen van het product kunnen worden geïnterpreteerd als interval-overlevingskansen (vergelijk de actuariële methode, zie volgende paragraaf). Omdat de Kaplan-Meier methode uit de actuariële methode afgeleid kan worden door het aantal leeftijdsintervallen tot oneindig te laten naderen, wordt deze methode ook wel de produkt limiet methode genoemd.

Paragraaf 2.3 De actuariële methode.

Hierbij wordt uitgegaan van een gegeven reeks tijdstippen, waarop een cohort onderling onafhankelijke, identieke individuen waargenomen wordt. Op ieder tijdstip is bekend hoeveel individuen van het cohort nog in leven zijn en daarmee ook hoeveel individuen gedurende het voorafgaande tijdsinterval overleden zijn.

De volgende (nieuwe) grootheden spelen hierbij een rol:

- i index met betrekking tot de waarnemingstijdstippen
- l_i het aantal overlevenden op tijdstip x_i ,
- d_i het aantal individuen, dat over het interval $[x_i, x_{i+1})$ sterft, $= l_i - l_{i+1}$
- p_i, q_i de voorwaardelijke overlevings- resp. sterftefractie tussen de leeftijden x_i en x_{i+1} .

Notatie: $p_{0i} = \prod (j=0..i-1) p_j$.

Van deze stochastische grootheden worden de (voorwaardelijke) aannemelijkheidsfuncties en de bijbehorende kansgenererende functies bepaald. Maximalisatie van deze aannemelijkheidsfuncties levert de meest aannemelijke schatters van de onbekende parameters op. De momenten van de verschillende schatters kunnen bepaald worden met behulp van de kansgenererende functies. Omdat de sterfteprocessen van de verschillende individuen onderling onafhankelijk en identiek verdeeld verondersteld worden, zijn de aantallen overlevenden en overledenen binomiaal c.q. multinomiaal verdeeld.

Het voorwaardelijke aantal overlevenden l_i .

De voorwaardelijke kansverdelingsfunctie hiervan is:

$$\Pr(l_i + 1 \mid l_i) = \frac{l_i!}{(l_i + 1)!} \cdot d_i \cdot q_i^{d_i} \cdot p_i^{l_i + 1}$$

resp.

$$\Pr(l_i \mid l_0) = \frac{l_0!}{l_i! (l_0 - l_i)!} p_0^{l_i} \cdot (1 - p_0)^{l_0 - l_i}$$

met kansgenererende functie:

$$G(l_i \mid l_0)(s) = \sum_{0 \leq l_i \leq l_0} \Pr(l_i \mid l_0) \cdot s^{l_i} = (1 - p_0 + p_0 s)^{l_0}$$

waaruit volgt:

$$E(l_i \mid l_0) = l_0 \cdot p_0$$

$$\text{var}(l_i \mid l_0) = l_0 \cdot p_0 \cdot (1 - p_0)$$

$$E(d_i \mid l_0) = l_0 \cdot p_0 \cdot q_i$$

$$\text{var}(d_i \mid l_0) = l_0 \cdot p_0 \cdot q_i \cdot (1 - p_0 \cdot q_i)$$

De gemeenschappelijke aantallen overlevenden en overledenen.

De gemeenschappelijke kansverdelingsfunctie luidt:

$$\Pr((l_i) \mid l_0) = \frac{l_0!}{l_i! (l_0 - l_i)!} q_i^{d_i} \cdot p_i^{l_i + 1}$$

Deze formule kan inductief naar het aantal leeftijdsintervallen bewezen worden. De bijbehorende kansgenererende functie luidt:

$$G((l_i) \mid l_0)(s) = (1 - \Pi(i) p_0 + \Pi(j < i) s_j + (1 - s_i))^{l_0}$$

waaruit volgt:

$$\text{cov}(l_i, l_j \mid l_0) = l_0 \cdot (1 - p_0) \cdot p_0 \quad (\text{igv } i < j)$$

$$\text{cov}(l_i, d_j \mid l_0) = l_0 \cdot (1 - p_0) \cdot p_0 \cdot q_j$$

$$\text{cov}(d_i, d_j \mid l_0) = - l_0 \cdot p_0 \cdot q_i \cdot p_0 \cdot q_j$$

De overlevings- en sterftekanssen.

De kansverdelingsfunctie van de optredende sterfte luidt:

$$\begin{aligned} \Pr(d_i) &= \text{cst} * \Pi(i) (p_{0i} * q_i)^{d_i} \\ &= \text{cst} * (1-p_1)^{d_0} * p_1^{l_1} * (1-p_2)^{d_1} * p_2^{l_2} * \dots \end{aligned}$$

Deze kansverdelingsfunctie bestaat uit een hiërarchische reeks conditionele kansverdelingsfuncties, die afzonderlijk gemaximaliseerd kunnen worden.

Hieruit volgt:

$$\begin{aligned} \hat{p}_i &= \frac{\sum_{j \geq i} \Pi_j}{\sum_{j \geq i} \Pi_j} = l_{i+1} / l_i \\ E(\hat{p}_i) &= E(E(l_{i+1} | l_i) / l_i) = p_i \\ \text{var}(\hat{p}_i) &= E(1/l_i) * p_i * (1-p_i) = p_i * (1-p_i) / E(l_i) \\ \text{cov}(\hat{p}_i, \hat{p}_j) &= 0 \text{ (igv } i \neq j) \end{aligned}$$

Paragraaf 2.4 De sarnemelijkheidafunctie met gebruik van de mortaliteitsdruk.

Een (parameter vrije) schatter van de leeftijdsafhankelijke mortaliteitsdruk kan afgeleid worden uit een schatter van de overlevingsfunctie:

$$\hat{H}(x) = - \ln \hat{S}(x)$$

De variantie van deze schatter wordt daarmee:

$$\text{var} \hat{H}(x) = \text{var} \hat{S}(x) / \hat{S}(x)^2$$

In het geval de (geparametriseerde) vorm van de mortaliteitsdruk of overlevingsfunctie als een gegeven wordt verondersteld kan de methode van de meeste sarnemelijkheid toegepast worden voor het schatten van de parameterwaarde(n). De zogenaamde meest sarnemelijke schatter(s) van de onbekende parameter(s) wordt gevormd door die parameterwaarde(n), die de sarnemelijkheidsfunctie maximaliseert.

Meestal wordt van de veronderstelling uitgegaan, dat de afzonderlijke individuen een identiek gedrag vertonen en onderling onafhankelijk zijn. Deze veronderstelling leidt ertoe, dat de sarnemelijkheidsfunctie geschreven kan worden als het produkt over de individuen van de (gelijkvormige) sarnemelijkheidsfuncties van de afzonderlijke overlijdenstijdstippen. Deze afzonderlijke sarnemelijkheidsfuncties kunnen geïnterpreteerd worden als de kans,

dat een individu overleeft tot zijn overlijdenstijdstip en dan sterft:

$$L = \prod_{i \in N} l_i(t_i) * h_i(t_i) = \prod_{i \in N} \exp^{-H_i(t_i)} * h_i(t_i)$$

met:

i index met betrekking tot de individuen

l_i de overlevingsfunctie van individu i

De logarithme $l = \ln L$ hiervan luidt:

$$l = \ln L = \sum_{i \in N} (\ln h_i(t_i) - H_i(t_i))$$

Wordt de mortaliteitsdruk gegeven in de vorm van een functievoorschrift met nog onbekende parameter b , dan levert maximalisatie van (de logarithme van) de aannemelijkheidsfunctie de meest aannemelijke schatter van b op. Twee voorbeelden voor cohorten, bestaande uit individuen met identieke overlevingsfuncties zijn:

Constante mortaliteitsdruk.

Onder aanname van een constante mortaliteitsdruk c voor alle individuen worden de logarithme van de aannemelijkheidsfunctie en de scoregelijkheid:

$$l(t_i|c) = \#(N) \ln c - c \sum_{i \in N} t_i$$

$$\frac{d}{dc} l(t_i|c) = \frac{\#(N)}{c} - \sum_{i \in N} t_i = \#(N) * \left(\frac{1}{c} - \frac{\sum_{i \in N} t_i}{\#(N)} \right) = 0$$

waaruit als meest aannemelijke schatter van de parameter c volgt:

$$\hat{c} = \#(N) / \sum_{i \in N} t_i$$

De grootte $\sum_{i \in N} t_i$ vormt de som van $\#(N)$ onderling onafhankelijke, exponentieel verdeelde stochasten en is dus gamma-verdeeld met parameters c en $\#(N)$. Dit geeft de mogelijkheid om direct hypothesen omtrent de factor c te toetsen.

De gegeven aannemelijkheidsfunctie komt overeen met die van $\sum_{i \in N} t_i$ trekkingen uit een Poisson-verdeling met som van de uitkomsten $\#(N)$. Deze overeenkomst vindt men ook terug in (bijvoorbeeld) de wachttijd-theorie: een Poisson-verdeeld aankomstproces van loket-klanten komt overeen met een exponentieel verdeeld 'interarrival time' van de klanten.

Weibull-verdeling van overleving.

Onder aanname van een Weibull-verdeling van de overleving:

$$h(x) = a * b * x^{b-1}$$

wordt de samenteljkheidsfunctie:

$$l(t_i|a,b) = \#(N) \ln ab + (b-1) \sum (t_i) - a \sum (t_i)^b$$

waaruit als meest samenteljk schatter van a voor gegeven b volgt:

$$\hat{a} = \#(N) / \sum (t_i)^b$$

Deze uitkomst vertoont een duidelijke overeenkomst met de meest samenteljk schatter van de parameter van een exponentiële overleving. Een Weibull-verdeeld overlijdenstijdstip T met parameter b geeft aan, dat het b -voudige overlijdenstijdstip T^b exponentieel verdeeld is.

Paragraaf 2.5 Type 1 censurering.

In de praktijk wordt van een groep individuen meestal slechts een beperkt aantal overlijdenstijdstippen waargenomen. Op het moment van afsluiten van de waarnemingsperiode is nog een aantal individuen in leven, waarvan dus de gerealiseerde overlijdenstijdstippen onbekend blijven. Het werkelijke overlijdenstijdstip is gecensureerd, maar valt in ieder geval na de beëindiging van de observatie. Deze beperking van de waarnemingen wordt censurering genoemd (Engels: censored survival analysis).

Verschillende vormen van censurering kunnen onderscheiden worden. De belangrijkste hiervan zijn: type 1 en type 2 censurering. Type 1 censurering houdt in: de sterfte van een individu wordt slechts waargenomen, indien deze plaatsvindt binnen een gegeven tijdsperiode. Deze tijdsperiode mag in principe van individu tot individu verschillen, maar wordt verondersteld niet van invloed te zijn op het overlijdenstijdstip. Anders gezegd: het censurerings-mechanisme is niet-informatief.

De volgende nieuwe grootheden worden nu toegepast:

- c_i de maximale waarnemingsduur van individu i (eventueel stochastisch),
- T_i de gerealiseerde waarnemingsduur van individu i ,
 $T_i = \min(t_i, c_i)$
- N de verzameling van individuen, waarvan de sterfte wordt waargenomen,

Het principe van type 1 censurering kan aan de hand van het volgende schema toegelicht worden:

individu

1 ----- t1
----- c1 T1 = c1

2 ----- t2
----- c2 T2 = t2

Van individu 1 wordt het overlijden niet, van individu 2 wel waargenomen.

Voor de drie genoemde parametervrije schattingsmethoden moeten de schattingsfuncties gecorrigeerd worden voor de ontbrekende overlijdenstijdstippen.

Bij de gereduceerde steekproef methode wordt voor ieder tijdstip de overlevingskans geschat uit het aantal overledenen tot dan gedeeld door de oorspronkelijke omvang van het cohort. In geval van censurering wordt de noemer gevormd door het aantal individuen, die óf op het betreffende tijdstip nog in leven zijn óf daarvoor gestorven zijn. De (gecensureerde) individuen, waarvan het censurerings-tijdstip valt vóór het betreffende tijdstip, worden buiten beschouwing gelaten.

Bij de Kaplan-Meier methode moet op ieder tijdstip de risicogroep bekend zijn. Voor censurerings-tijdstippen, die samenvallen met overlijdenstijdstippen, wordt aangenomen, dat de censurering net iets later heeft plaatsgevonden. Dus een individu, dat op een tijdstip gecensureerd wordt, behoort op dat moment nog tot de risicogroep.

Bij de actuariële methode is slechts bekend tot welk tijdsinterval censurerings-tijdstippen behoren. Het precieze censurerings-tijdstip is echter onbekend. Een mogelijkheid is om te veronderstellen, dat (in verwachting) de censurerings-tijdstippen halverwege een tijdsinterval vallen.

In geval van parametrische methoden dient de aannemelijkheidsfunctie ook termen te bevatten voor de censurerings-tijdstippen. De aannemelijkheidsfunctie L_1 van de gerealiseerde waarnemingsduren onder type 1 censurering kan als volgt afgeleid worden:

$$\frac{d}{dx} \text{Pri}(x \leq x \mid x \leq c_i) = h_i(x) * l_i(c_i) / 1 - l_i(c_i)$$

$$\text{Pr}(M) = \frac{\#(N)!}{\#(N) \cdot \#(N-M)!} \prod_{i \in M} l_i(c_i) \prod_{i \in N-M} 1 - l_i(t_i)$$

$$\begin{aligned} L_1 &= \text{cst} * \prod_{i \in M} \exp^{-H_i(t_i)} h_i(t_i) * \prod_{i \in N-M} \exp^{-H_i(c_i)} \\ &= \text{cst} * \prod_{i \in N} \exp^{-H_i(T_i)} * \prod_{i \in M} h_i(T_i) \end{aligned}$$

Paragraaf 2.6 Type 2 censurering.

Type 2 censurering houdt in: slechts de eerste m sterfgevallen worden geregistreerd. De volgende grootheden worden nu ingevoerd:

- m het gegeven aantal individuen, waarvan de sterfte waargenomen wordt,
- r het gegeven aantal individuen, waarvan de sterfte niet waargenomen wordt,
- $t(m)$ de stochastische algemene waarnemingsperiode,

Het principe van type 2 censurering kan aan de hand van het volgende schema toegelicht worden:

individu		het einde van de waarnemingsperiode
1	-----	
2	-----	
3	-----	

De eerste twee sterfgevallen worden slechts waargenomen.

Ook type 2 censurering heeft consequenties voor de formuleringen van de parameter vrije schattingsfuncties. In het algemeen kan gesteld worden, dat de genoemde parameter vrije schatters onder type 2 censurering identiek zijn aan de niet-gecensureerde schatters tot aan waarnemingsduur $t(m)$, maar voor tijdstippen daarna niet meer gedefinieerd zijn.

In geval van parametrische schattingsmethode wordt de aannemelijkheidsfunctie L_2 van de gerealiseerde levensduren onder type 2 censurering:

$$L_2 = cst * \prod_{i \in M} \exp^{-H_i(t_i)} h_i(t_i) * \prod_{i \in N-M} \exp^{-H_i(t(m))}$$

Hieruit blijkt, dat indien voor type 1 censurering de waarnemingsperiode van alle individuen $c_i = t(m)$ gekozen wordt de aannemelijkheidsfuncties L_1 en L_2 identiek zijn. Type 2 censurering kan dus worden beschouwd als een speciaal geval van type 1 censurering.

HOOFDSTUK 3 De partiële aannemelijkheidsfunctie.

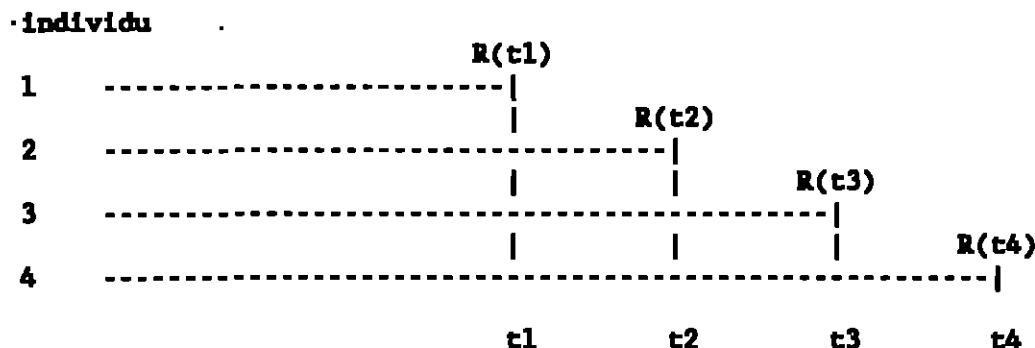
Paragraaf 3.1 Opplitsing van de volledige aannemelijkheidsfunctie.

De aannemelijkheidsfunctie zonder censurering beschrijft de waarschijnlijkheid van optreden (realisatie) van het overlijdens-tijdstip van een aantal individuen. Een andere benadering van de optredende sterfte leidt tot een herformulering van deze aannemelijkheidsfunctie. Hierbij wordt voor ieder tijdstip van overlijden de waarschijnlijkheid gegeven, dat juist het betreffende individu overlijdt. Deze waarschijnlijkheid is de verhouding tussen de sterftekansen van het betreffende individu en die van alle op dat moment nog in leven zijnde individuen (inclusief de betreffende). Deze alternatieve benadering leidt tot de partiële aannemelijkheidsfunctie, die met name zinvol is bij het schatten van de parameter(s) van de mortaliteitsdruk voor een heterogene populatie (zie ook hoofdstuk 4).

De volgende (nieuwe) grootheden worden geïntroduceerd:

- t_i het i -de tijdstip van overlijden. Vergelijk met oorspronkelijke betekenis: het tijdstip van overlijden van individu i ,
- $R(x)$ alle individuen, die op tijdstip x nog leven, ofwel: de risicoverzameling op tijdstip x , $= \{ i \in N \mid t_i \geq x \}$

De verzamelingen $R(t_i)$ vormen een geneste reeks. Het volgende voorbeeld laat een toepassing van deze grootheid zien:



Met behulp van de verzamelingen $R(t_i)$ kan de aannemelijkheidsfunctie van de overlijdenstijdstippen herschreven worden als:

$$L = \prod_i e^{-H_1(t_i)} * h_1(t_i) \\ = \prod_i j_i \prod_i(t_i) \exp^{-\int_{t_{i-1}}^t h_j(s) ds} * h_j(t_i) * \quad (L1)$$

$$\prod_i \sum_{j \in R(t_i)} \frac{h_1(t_i)}{h_j(t_i)} h_j(t_i) \quad (L2)$$

De aannemelijkheidsfunctie kan in twee produkttermen uitgesplitst worden. De term L1 beschrijft de kansverdeling van de overlijdenstijdstippen, de term L2 de kansverdeling van de selectie van gestorvenen. L2 wordt de partiële aannemelijkheidsfunctie genoemd (zie bijvoorbeeld Oakes, 1981).

De partiële aannemelijkheidsfunctie blijkt, afgezien van eventuele constanten, identiek te zijn aan de volledige aannemelijkheidsfunctie, indien voor ieder tijdsinterval $[t_{i-1}, t_i)$ voor ieder nog levend individu j de mortaliteitsdruk kan worden geschreven als:

$$h_j(x|b) = z_j(b) * h_{Ci} \quad (t_{i-1} \leftarrow x < t_i)$$

In dit geval vormen de mortaliteitsdrukken het produkt van een individuele risicofactor z_j en een collectieve mortaliteitsdruk h_{Ci} (zie ook paragraaf 4.5). Beide zijn afhankelijk van, maar constant over ieder tijdsinterval. In dit geval reduceert de partiële aannemelijkheidsfunctie tot een functie, waarin slechts de individuele risicofactoren voorkomen en niet de collectieve mortaliteitsdruk. In de volgende paragrafen wordt de relatie tussen partiële en volledige aannemelijkheidsfunctie verder uitgewerkt.

Paragraaf 3.2 De geneste structuur van de gebeurtenissen.

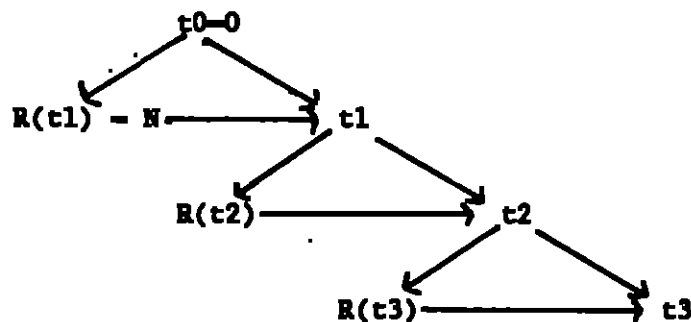
De genoemde twee termen van de aannemelijkheidsfunctie vormen twee geneste structuren. Deze nesting treedt zowel binnen als tussen elk van beide termen op. Voor beide termen zal deze nesting eerst verbaal beschreven worden, vervolgens formeel. Op basis hiervan worden de samenstellende kansdichtheidsfuncties van beide termen herschreven.

De afzonderlijke kansdichtheidsfuncties van de overlijdenstijdstippen zijn voorwaardelijk met betrekking tot de voorafgaande overlijdenstijdstippen en de individuen, die op het moment nog in leven zijn, t_i resp. $R(t_i)$. Deze conditionerende gebeurtenissen kunnen op een geneste wijze gestructureerd worden:

t_i hangt af van t_{i-1} en $R(t_i)$

$R(t_i)$ hangt af van t_{i-1}

of in schematische vorm:



Paragraaf 3.3 De aannemelijkheidsfunctie geherformuleerd.

De i -de kansdichtheidsfunctie van de produktterm $L1$ van de aannemelijkheidsfunctie kan nu als volgt geschreven worden:

$$f_i(t | t_{i-1}, R(t_i)) = \frac{d}{dt} \exp^{-\sum_{j=1}^i R(t_j)} \int_{t_{i-1}}^{t_i} h_j(s) ds$$

Dit is een continue kansdichtheidsfunctie. De volledige eerste produktterm van de aannemelijkheidsfunctie wordt hiermee:

$$L1 = \prod_{i \in N} f_i(t_i | t_{i-1}, R(t_i))$$

De i -de kansdichtheidsfunctie van de tweede produktterm $L2$, de partiële aannemelijkheidsfunctie, wordt:

$$g_i(j | R(t_i)) = \frac{h_j(t_i)}{\sum_{k=1}^K h_k(t_i) - h_k(t_{i-1})}$$

Dit is een discrete kansverdelingsfunctie. De volledige tweede term wordt:

$$L2 = \prod_{i \in N} g_i(R(t_{i+1}) - R(t_i) | t_i)$$

De functies $L1$ en $L2$ bestaan beide uit een reeks gelijkvormige kansdichtheids- resp. kansverdelingsfuncties.

Paragraaf 3.4 De eerste term van de aannemelijkheidsfunctie.

De eerste produktterm $L1$ van de aannemelijkheidsfunctie kan met behulp van de volgende grootheid herschreven worden:

$$M'(x) = \Pi(i \in N) h_i(x) / \#(N)$$

$$M'(x) = \Pi(i \in R(x)) h_i(x) / M'(x)$$

$M'(x)$ geeft de ongewogen gemiddelde mortaliteitsdruk weer, $M'(x)$ een gecorrigeerd aantal individuen. De produktterm L_1 wordt:

$$L_1 = \exp^{-\int_0^\infty M'(s) M'(s) ds} * \prod_{i \in N} M'(t_i) M'(t_i)$$

Deze functie vormt de aannemelijkheidsfunctie van de overlijdenstijdstippen van een aantal identieke individuen. Deze herformulering van L_1 suggereert, dat het onderscheid tussen de individuen met name door de partiële aannemelijkheidsfunctie beschreven wordt.

Paragraaf 3.5 Eigenschappen van de partiële aannemelijkheidsfunctie.

De partiële aannemelijkheidsfunctie beschrijft het onderscheid tussen de oorspronkelijke individuen en bestaat uit het produkt van een reeks gelijkvormige kansdichtheidsfuncties. Deze kansdichtheidsfuncties beschrijven, gegeven de risicoverzameling op een overlijdenstijdstip, de kans op het overlijden van een specifiek individu.

De partiële aannemelijkheidsfunctie is formeel geen aannemelijkheidsfunctie, omdat de overlijdenstijdstippen als een conditionale in plaats van stochastische gebeurtenis beschouwd worden. De afzonderlijke samenstellende kansdichtheidsfuncties zijn het echter wel. De partiële aannemelijkheidsfunctie als produkt van de samenstellende kansdichtheidsfuncties voldoet daarom aan een aantal eigenschappen van aannemelijkheidsfuncties. Zij:

b de te schatten parameter van de mortaliteitsdruk

$D = \frac{d}{db}$ de differentiaal-operator met betrekking tot b

Dan kan afgeleid worden:

De logaritme van de partiële aannemelijkheidsfunctie.

$$l = \ln L_2 = \sum(i) \ln g_i(j | R(t_i))$$

$$= \sum(i) (\ln h_j(t_i) - \ln \sum(k \in R(t_i)) h_k(t_i))$$

Beschouw nu een willekeurige kansdichtheidsfunctie g_i met $j = R(t_i+1) - R(t_i)$, dan worden de verwachting en variantie van de scorefunctie:

De verwachting:

$$\begin{aligned} E(D \ln g_i(j \mid R(t_i))) \\ = \sum(j) (D \ln g_i(j \mid R(t_i)) * g_i(j \mid R(t_i))) \\ = \sum(j) (\frac{D h_j(t_i)}{h_j(t_i)} - \frac{\sum(k) D h_k(t_i)}{\sum(k) h_k(t_i)}) * \frac{h_j(t_i)}{\sum(k) h_k(t_i)} = 0 \end{aligned}$$

De variantie:

$$\begin{aligned} E(DD' \ln g_i(j \mid R(t_i))) \\ = \sum(j) (DD' \ln g_i(j \mid R(t_i)) * g_i(j \mid R(t_i))) \\ = - \text{var}(D \ln g_i(j \mid R(t_i))) \end{aligned}$$

Onderlinge onafhankelijkheid van de scorefuncties (11 < 12):

$$E(D \ln g_{i1}(j \mid R(t_{i1}))' D \ln g_{i2}(j \mid R(t_{i2}))) = 0$$

Vervolgens wordt verondersteld, dat de wet van de grote aantallen toepasbaar is. In dat geval en uit de bovenbeschreven eigenschappen volgt, dat de meest aannemelijke schatter op basis van de partiële aannemelijkheidsfunctie asymptotisch zuiver is. Deze schatter is bovendien asymptotisch normaal verdeeld met variantie gelijk aan de inverse van de Fisher informatie-matrix $I = - DD' \ln L_2$.

Een strikt bewijs van de asymptotische eigenschappen van de partiële aannemelijkheidsfunctie kan gegeven worden met behulp van de theorie van tellingen en martingalen (zie volgende paragraaf).

Paragraaf 3.6 Tellingen en martingaal-theorie.

Het nieuwe kader, waarin het bovenstaande geplaatst wordt, wordt geleverd door de theorie van tellingen en martingalen en kan als volgt beschreven worden (zie Gill, 1984). Een cohort wordt gedurende een tijdsperiode gevolgd, waarbij onttrekking plaatsvindt ten gevolge van sterfte of censurering. Beide, sterfte en censurering, kunnen op het niveau van een individu worden beschreven als de eenheidsstap (Engels: jump) van een indicatorfunctie, met andere woorden in termen van tellingen. Het sterfteproces van het cohort vormt aldus een multivariaat tellingen-proces (Engels: counting process).

Daarnaast kan sterfte beschreven worden in termen van de aggregatie van risico en van de overlevingscurve van een cohort (zie ook hoofdstuk 4). Het risico betreft de niet-observeerbare, geaggregeerde mortaliteitsdruk, de (logaritme van de) overlevingscurve beschrijft het optreden van sterfte. Vervacht mag worden, dat in verwachting beide grootheden gelijk zijn. Het

verschil tussen beide wordt een martingaal genoemd. Een martingaal is vergelijkbaar met witte ruis (Engels: white noise).

Voor een nette beschrijving van het genoemde tellingen-proces en de martingaal wordt van de volgende nieuwe grootheden gebruik gemaakt (zie ook paragraaf 2.1 en 2.3):

$$n_1(x) = I(T_1 \leq x, T_1 \leq c_1) \quad \text{indicatorfunctie}$$

$$n = (N_1(x) \mid x \geq 0, i \in N)$$

$$y_1(x) = I(T_1 > x, c_1 > x) \quad \text{indicatorfunctie}$$

$$h_1(x) \quad \text{stochastische mortaliteitsdruk}$$

Voor ieder individu i beschrijft $n_i(\cdot)$ het optreden van een gebeurtenis (sterfte of censurering) en $y(\cdot)$ de observatieperiode.

$n_i(\cdot)$ vormt een tellingen-proces:

- $n_i(\cdot)$ vormt een stapfunctie met stapgrootte (jump) 1,
- $n_i(0) = 0$,
- $n_i(\cdot)$ is rechts-continu.

$n(\cdot)$ vormt een multivariaat tellingen-proces:

- De functies $n_i(\cdot)$ vormen tellingen-processen,
- Geen twee gebeurtenissen (jumps) vinden tegelijkertijd plaats.

Verondersteld wordt, dat de individuele mortaliteitsdruk $h_i(x)$ het produkt vormt van een individuele (risico-)factor en een collectieve mortaliteitsdruk (zie ook paragraaf 4.5). De eerste term wordt (gegeneraliseerd) lineair bepaald door de stochastische risicovariabele z_i :

$$h_1(t) = y_1(t) * f(z_1' b) * h_C(t)$$

De term $y_i(t)$ is toegevoegd om aan te geven of een individu nog niet onttrokken is aan de waarneming. De genoemde grootheden voldoen aan de volgende eigenschappen:

- n_i , y_i en z_i kunnen gemeten worden (Engels: are observable);
- y_i en z_i zijn voorspelbaar (Engels: are predictable): $y_i(t)$ en $z_i(t)$ zijn vastgelegd, gegeven het verleden tot aan tijdstip t , aangeduid met $F(t-)$.

De oorspronkelijke volledige samenteljkheidsfunctie kan nu worden herschreven tot:

$$L(t) = \prod_{i \in N} \prod_{s \leq t} \left(\frac{y_i(s) f(z_i' b)}{1 - \frac{y_i(s) f(z_i' b)}{1 - \frac{y_i(s) f(z_i' b)}{1 - \dots}}} \right) dN_i(s)$$

Deze aannemelijkheidsfunctie lijkt op de oorspronkelijke partiële aannemelijkheidsfunctie, maar beide verschillen als volgt:

- In de noemer van de breuk is de risicoverzameling vervangen door de volledige verzameling,
- Een factor met betrekking tot de observatie is toegevoegd,
- De overlijdenstijdstippen zijn (indirect) via de tellingenprocessen toegevoegd.

Op basis van de genoemde grootheden kan de volgende reeks functies geconstrueerd worden:

$$M_1(t) = m_1(t) - \int_0^t h_1(s) ds$$

$M_1(t)$ vormt een martingaal, omdat:

$$\begin{aligned} E(dM_1(t) \mid F(t-)) &= E(dm_1(t) - h_1(t)dt \mid F(t-)) \\ &= E(dm_1(t) \mid F(t-)) - E(h_1(t)dt \mid F(t-)) \\ &= \Pr(dm_1(t)=1 \mid F(t-)) - h_1(t) dt - h_1(t) dt - h_1(t) dt \\ &= 0 \end{aligned}$$

De variantie $\langle M_1 \rangle(t)$ (Engels: predictable variation process) van $M_1(t)$ volgt uit:

$$\begin{aligned} d\langle M_1 \rangle(t) &= \text{var}(dM_1(t) \mid F(t-)) = E(dM_1(t)^2 \mid F(t-)) \\ &= E((dm_1(t) - h_1(t) dt)^2 \mid F(t-)) \\ &= E((dm_1(t) - E(dm_1(t)))^2 \mid F(t-)) \\ &= h_1(t) dt + (1 - h_1(t) dt) \rightarrow h_1(t) dt \end{aligned}$$

De covariantie $\langle M_1, M_j \rangle$ (Engels: predictable covariation process) van $M_1(.)$ en $M_j(.)$ volgt uit:

$$\begin{aligned} d\langle M_1(t), M_j(t) \rangle &= \text{cov}(dM_1(t), dM_j(t) \mid F(t-)) \\ &= E(dM_1(t) * dM_j(t) \mid F(t-)) - E(dM_1(t) \mid F(t-)) * E(dM_j(t) \mid F(t-)) \\ &= E((dm_1(t) - h_1(t)dt) * (dm_j(t) - h_j(t)dt) \mid F(t-)) \\ &= 0 - h_1(t)dt * 0 - h_j(t)dt * 0 + h_1(t)dt * h_j(t)dt \rightarrow 0 \end{aligned}$$

De scorefunctie van de beschreven aannemelijkheidsfunctie wordt:

$$\begin{aligned}
M^{(n)}(t) &= D \ln L(t) / \sqrt{n} = \sum_{i=1}^n \sum_{u \leq t} (x_i(u) - G(u)) dm_i(u) / \sqrt{n} \\
&= \sum_{i=1}^n \int_{u \leq t} (x_i(u) - G(u)) dm_i(u) / \sqrt{n} \\
&= \sum_{i=1}^n \int_{u \leq t} (x_i(u) - G(u)) / \sqrt{n} dm_i(u)
\end{aligned}$$

$$\text{met: } G(u) = \frac{\sum_{j=1}^p y_j(u) x_j(u) f'(x_j' b)}{\sum_{j=1}^p y_j(u) f'(x_j' b)}$$

Ieder element van de somreeks $M(n)$ bestaat uit de convolutie van een voorspelbare stochast en een martingaal. Deze eindige somreeks vormt eveneens een martingaal. Dit geeft de mogelijkheid om de centrale limietstelling van martingalen op $M(n)$ toe te passen. Onder de veronderstellingen (waaraan in dit geval voldaan wordt), dat in de limiet:

- de sprongen (jumps) van $M(n)$ kleiner worden en $M(n)$ aldus continu is,
- de variantie (predictable variation process) $\langle M(n) \rangle$ deterministisch wordt,

en de (bovenbewezen) eigenschap, dat de covarianties van $M(n)$ onderling onafhankelijk zijn, is $M(n)$ asymptotisch normaal verdeeld met verwachting 0 en variantie gelijk aan de deterministische limiet van $\langle M(n) \rangle$.

Het bovenstaande geeft aan, dat de partiële aannemelijkheidsfunctie een nette aannemelijkheidsfunctie vormt, waarmee de meest aannemelijke schatter van de onbekende parameter b bepaald kan worden.

HOOFDSTUK 4 Heterogeniteit.

Paragraaf 4.1 Inleiding.

In hoofdstuk 1 is de overlevingsfunctie geïntroduceerd. Deze kan geïnterpreteerd worden in termen van de overlevingscurve van een geboorte-cohort en als het complement van de kansverdelingsfunctie van het overlijdenstijdstip van een willekeurig individu. Hierbij werd impliciet verondersteld, dat het geboorte-cohort bestaat uit een verzameling onderling onafhankelijke identieke individuen (homogeen cohort, zie paragraaf 1.1). In werkelijkheid wordt aan deze veronderstelling niet voldaan en bestaat ieder cohort uit verschillende en eventueel onderling afhankelijke individuen (heterogeen cohort).

In paragraaf 4.4 wordt getoond, dat het echter onmogelijk is om slechts op basis van individuele overlijdenstijdstippen te bepalen of een populatie homogeen of heterogeen is. Daartoe dienen meer gegevens verzameld te worden. Dit kan in het algemeen op twee manieren gebeuren. Ten eerste: van de individuen wordt naast de overlijdenstijdstippen tevens een aantal risicovariabelen gemeten. Ten tweede: de gemeten individuele overlijdenstijdstippen worden gebruikt om uitspraken te doen over een (mogelijk) onderscheid tussen subpopulaties. In statistisch-epidemiologische termen: in het eerste geval zijn experimentele en observationele eenheid gelijk (individu), in het tweede geval zijn beide verschillend (subpopulatie resp. individu).

Het onderscheid tussen beide manieren kan ook beschreven worden in termen van observeerbaarheid. In het eerste geval zijn de risicofactoren te meten (observeerbaar). In het tweede geval dient het (subpopulatie-)risico indirect uit de gemeten overlijdenstijdstippen bepaald te worden (niet-observeerbaar). In hoofdstuk 5 wordt aan het onderwerp risicovariabelen meer aandacht besteed.

Paragraaf 4.2 Heterogeniteit en selectie.

De overlevingscurve beschrijft voor iedere leeftijd de fractie van het oorspronkelijke geboorte-cohort, dat nog in leven is. De relatieve verandering van deze fractie kan geassocieerd worden met de mortaliteitsdruk. Deze mortaliteitsdruk heeft slechts betrekking op individuen, die nog in leven zijn.

Daarnaast kunnen de overlevingsfuncties van alle individuen van het geboorte-cohort samengevoegd worden tot een gemiddelde overlevingsfunctie. De relatieve verandering van deze gemiddelde overlevingsfunctie wordt eveneens beschreven door de mortaliteitsdruk. Deze mortaliteitsdruk heeft betrekking op alle individuen en wijkt in het algemeen af van de mortaliteitsdruk op basis van de overlevingscurve (zie vorige alinea). Slechts indien alle individuen identiek zijn, zijn beide mortaliteitsdrukken gelijk. In dat geval wordt de bevolking homogeen genoemd.

De overlevingsfunctie en daarmee de mortaliteitsdruk van een individu kunnen niet gemeten worden, slechts de overlevingsfunctie (-curve) van een (geboorte-)cohort. Op basis van de overlevingsfunctie van een cohort kan geprobeerd worden om uitspraken te doen over die van de afzonderlijke individuen van het cohort. Hierbij bezit men echter een zekere vrijheid. Bijvoorbeeld: in het geval van een dalende mortaliteitsdruk kan men twee extreme gevallen onderscheiden. Ten eerste: de individuele mortaliteitsdruk is constant over de tijd, maar verschilt over de individuen (heterogene populatie). De gemiddelde mortaliteitsdruk neemt af op grond van selectie. De zwakke individuen met een hoge mortaliteitsdruk overlijden (in verwachting) het eerst, waardoor het relatief gezonde deel van de bevolking overblijft. Ten tweede: de individuele mortaliteitsdruk neemt af in de loop van de tijd, maar is voor ieder tijdstip constant over alle individuen (homogene populatie). De collectieve mortaliteitsdruk komt overeen met de (iedere) individuele mortaliteitsdruk.

Paragraaf 4.3 Formele introductie.

In deze paragraaf wordt heterogeniteit van een cohort formeel geïntroduceerd. De heterogeniteit van de bevolking wordt weergegeven door de aanwezigheid van verschillende individuele overlevingsfuncties. Bij de geboorte wordt aan ieder individu willekeurig een mortaliteitsdruk toegekend, die de kansverdeling van zijn (stochastisch) overlijdenstijdstip beschrijft.

Bij de formele beschrijving van de toekening van individuele overlevingsfuncties wordt van de volgende (nieuwe) grootheden gebruik gemaakt:

- Z de verzameling van mogelijke toekenningen z . Iedere toekening representeert een mortaliteitsdruk,
- $f(z)$ de kansdichtheidsfunctie van z ,
- $h_z(t)$ de mortaliteitsdruk op leeftijd t voor toekening z ,

$h_C(t)$ de over Z gemiddelde mortaliteitsdruk (Engels: basic intensity),

$S(t)$ de overlevingscurve van het cohort,

$h(t)$ de mortaliteitsdruk, 'volgend uit de overlevingscurve, (Engels: population intensity).

Van de genoemde mortaliteitsdrukken kan slechts $h(t)$ geobserveerd worden. Deze groothed beschrijft (in verwachting) de overlevingscurve:

$$\exp^{-\int_0^t h(s) ds} = E \left(\exp^{-\int_0^t h_z(s) ds} \right)$$

Hierbij wordt de verwachting E van de toekennings-variabele z genomen over de verzameling Z . Uitwerking van deze formule levert op:

$$h(t) = \frac{\int_Z h_z(t) \exp^{-\int_0^t h_z(s) ds} f(z) dz}{\int_Z \exp^{-\int_0^t h_z(s) ds} f(z) dz}$$

Hieruit blijkt, dat de mortaliteitsdruk $h(t)$ geschreven kan worden als een gewogen gemiddelde van de individuele mortaliteitsdrukken. De wegingscoëfficiënten worden bepaald door de overlevingsfunctie voor iedere toekenning en de kansverdeling van deze toekenningen over het cohort. De individuele overlevingsfuncties worden vastgelegd door de toegekende mortaliteitsdruk.

De gewogen gemiddelde mortaliteitsdruk wijkt af van de ongewogen gemiddelde mortaliteitsdruk:

$$h_C(t) = \int_Z h_z(t) f(z) dz = E(h_z(t))$$

Het ongewogen gemiddelde $h_C(t)$ is groter dan het gewogen gemiddelde $h(t)$. De interpretatie hiervan ligt voor de hand. Zwakke individuen met een hoge mortaliteitsdruk sterven (in verwachting) op relatief lage leeftijd en beïnvloeden daarom voor hoge leeftijden de gewogen gemiddelde mortaliteitsdruk in mindere mate dan de sterke individuen met een lage mortaliteitsdruk. Met andere woorden: wordt een heterogeen cohort (ten onrechte) homogeen verondersteld, dan wordt de individuele mortaliteitsdruk onderschat.

Analoog kan heterogeniteit gedefinieerd worden in geval van onderscheiden doodsoorzaken. Hierbij wordt gebruik gemaakt van de volgende (nieuwe) grootheden:

$h_{jz}(t)$ de mortaliteitsdruk voor doodsoorzaak J op leeftijd t voor toekenning z . $h_z(t) = \sum(j) h_{jz}(t)$.

$h_j(t)$ de verwachting van de mortaliteitsdruk voor
doodsoorzaak j op leeftijd t over de nog levende
individueen

De grootheid $h_j(t)$ kan geschreven worden als:

$$h_j(t) = \frac{\int_Z h_{jz}(t) \exp^{-\int_0^t h_z(s) ds} f(z) dz}{\int_Z \exp^{-\int_0^t h_z(s) ds} f(z) dz}$$

Hieruit volgt:

$$\sum(j) h_j(t) = h(t)$$

Paragraaf 4.4 Een voorbeeld van homogeniteit of heterogeniteit.

In deze paragraaf wordt een voorbeeld gegeven van de keuze-mogelijkheid van de toekenning van stochastische mortaliteitsdrukken op basis van een gemeten gemiddelde mortaliteitsdruk van een cohort. Twee extreme keuzes kunnen gemaakt worden: enerzijds een homogene bevolking met toenemende levensverwachting, anderzijds een heterogene bevolking met constante levensverwachting.

Gegeven is de gemiddelde mortaliteitsdruk:

$$h(t) = \frac{1}{1+t}$$

Deze mortaliteitsdruk kan het gevolg zijn van een:

Homogeen cohort

Ieder individu krijgt met kans 1 de mortaliteitsdruk $1/(1+t)$ toegewezen.

Heterogeen cohort

Ieder individu krijgt een constante mortaliteitsdruk r toegewezen, waarbij r negatief exponentieel verdeeld is met parameter 1:

$$\begin{aligned} h(t) &= - \frac{d}{dt} \ln E \left(\exp^{-\int_0^t h_z(s) ds} \right) = - \frac{d}{dt} \ln \int_0^\infty e^{-zs} e^{-z} dz \\ &= - \frac{d}{dt} \ln \frac{1}{1+t} = \frac{1}{1+t} \end{aligned}$$

Beide beschrijvingen leiden tot de gegeven gemiddelde mortaliteitsdruk.

Paragraaf 4.3 Multiplicatieve heterogeniteit.

In geval van niet-observeerbare grootheden kan de individuele mortaliteitsdruk beschouwd worden als het produkt van de collectieve cohort-specifieke mortaliteitsdruk en een (subpopulatie-specifieke, multiplicatieve) (risico-)variabele (Engels: *mixing* variabele, *frailty*). Deze variabele is niet direct te meten, maar wordt indirect uit de (sterfte-)gegevens afgeleid. In deze en de volgende paragrafen wordt aan deze vorm van heterogeniteit aandacht besteed.

Hierbij wordt gebruik gemaakt van de notatie, ingevoerd in paragraaf 4.3, en de volgende nieuwe notaties:

z de multiplicatieve (risico-)variabele met betrekking tot een individuele mortaliteitsdruk

De individuele mortaliteitsdruk vormt het produkt van de collectieve mortaliteitsdruk en de individu-specifieke tijdsafhankelijke risicovariabele:

$$h_z(t) = z * h_C(t)$$

De factor z is een niet-negatieve stochastische variabele, waarvan de verwachting over het cohort gelijk aan 1 is. De overlevingscurve wordt als volgt uitgedrukt in de mortaliteitsdrukken van de individuen van het cohort:

$$\begin{aligned} S(t) &= E(e^{-H_z(t)}) = E(e^{-z * H_C(t)}) \\ &= \int_0^\infty e^{-z * H_C(t)} f(z) dz = L(H_C(t)) \end{aligned}$$

De overlevingscurve van het cohort vormt de Laplace-getransformeerde van de collectieve mortaliteitsdruk. De kansdichtheidsfuncties van de schalingsfactor voor de gestorvenen resp. overlevenden op tijdstip t zijn:

$$\begin{aligned} f(z | T=t) &= \frac{f(z) h_z(t) e^{-H_z(t)}}{\int_0^\infty f(z) h_z(t) e^{-H_z(t)} dz} = \frac{f(z) z e^{-z * H_C(t)}}{L(H_C(t))} \\ f(z | D>t) &= \frac{f(z) e^{-z * H_C(t)}}{L(H_C(t))} \end{aligned}$$

met verwachting:

$$E_{t+}(z) = E(z | D>t) = - \frac{d}{dt} \ln L(H_C(t))$$

Paragraaf 4.6 Voorbeelden van multiplicatieve factoren.

In deze paragraaf worden twee specifieke keuzes voor de factor z beschreven.

Exponentiële families (Hougaard, 1984):

$$f_{a,b}(z) = z^a e^{-bz} m(z) / c(a,b)$$

De constante $c(a,b)$ voldoet aan:

$$(\text{kansdichtheidsfunctie:}) \quad c(a,b) = \int_0^\infty z^a e^{-bz} m(z) dz$$

$$(\text{verwachting 1:}) \quad c(a+1,b) = c(a,b)$$

De overlevingscurve van het cohort wordt in dit geval:

$$S(t) = c(a, b+H_G(t))$$

De voorwaardelijke kansdichtheidsfuncties van de risicovariabele worden:

$$f_{a,b}(z | T=t) = f_{a+1, b+H_G(t)}(z) / c(a,b)$$

$$f_{a,b}(z | T>t) = f_{a, b+H_G(t)}(z) / c(a,b)$$

Het blijkt, dat de kansdichtheidsfuncties van de risicovariabele onder de gestorvenen en overlevenden dezelfde vorm bezitten als de oorspronkelijke.

Als voorbeeld van een exponentiële familie wordt de gamma-verdeling beschouwd:

$$f_{a,n}(z) = \frac{a^n z^{n-1} e^{-az}}{\Gamma(n) \Gamma(a)^n}$$

met $a=n$ voor verwachting van z gelijk aan 1. Dan wordt:

$$S(t) = L(H_G(t)) = \left(\frac{\Gamma(a) \Gamma(a)}{\Gamma(a+H_G(t))} \right)^n$$

$$E_{t+}(z) = \frac{\Gamma(a) \Gamma(a)}{\Gamma(a+H_G(t))}$$

Voor de gamma-verdeling kan de mortaliteitsdruk op basis van de overlevingscurve (population intensity) beschreven worden in termen van de ongewogen collectieve mortaliteitsdruk (basic intensity):

$$h(t) = - \frac{d}{dt} \ln S(t) = \frac{h_G(t)}{1 + H_G(t) / a}$$

Dit verband leidt tot het tweede voorbeeld van heterogeniteit:

Generaliseerde gamma-verdeling (Aalen, 1988):

$$h(t) = \frac{h_G(t)}{(1 + a H_G(t) / b)^b}$$

Deze relatie tussen population en basic intensity kan voor $b > 1$ afgeleid worden, uit de veronderstelling, dat de schalingsfactor z samengesteld Poisson-verdeeld is:

$$z = \sum_{i=1..N} x_i \text{ i.g.v. } N > 0$$

$$= 0$$

$$N=0$$

met $\{x_i\}$ onderling onafhankelijk gamma(a,b)-verdeeld en N Poisson(r)-verdeeld. Hieruit volgt:

$$S(t) = L(H_G(t)) = e^{-r + r \left(\frac{b}{b + H_G(t)} \right)^a}$$

Differentiëren van deze formule en herdefiniëring van de constanten (met $b = r \cdot a$, omdat $E(z)=1$) levert de bovengenoemde relatie tussen population en basic intensity op. Deze relatie kan worden herschreven tot:

$$\frac{h(t)}{h_G(t)} = \left(1 + a H_G(t) / b \right)^{-b} \leftarrow 1$$

Deze verhouding neemt af voor toenemende grootte van de coëfficiënten a en b en risico $H_G(t)$.

Paragraaf 4.7 Heterogeniteit en partiële samenlijkhed.

De in hoofdstuk 3 geïntroduceerde partiële samenlijkhedsfunctie blijkt met name bruikbaar te zijn bij parameterschattingen voor een heterogeen veronderstelde populatie. Hierbij wordt verondersteld, dat slechts de multiplicatieve risicovariabele van de onbekende parameter afhangt:

$$h_{z|b}(t) = z(b) * h_G(t)$$

Dan worden de partiële samenlijkhedsfunctie L en de scoregelykheid:

$$\ln L = \sum(i) \ln z_i(b) - \sum(i) \ln \sum(j \in R(t_i)) z_j(b)$$

$$D \ln L = \sum(i) \frac{D z_i(b)}{z_i(b)} - \sum(i) \frac{\sum(j \in R(t_i)) D z_j(b)}{\sum(j \in R(t_i)) z_j(b)} = 0$$

Het blijkt, dat de basic intensity h_G niet meer in de scorefunctie en daarmee in de meest samenlijke schatter van b voorkomt. Wordt bovendien een proportioneel (multiplicatief) risico verondersteld:

$$z_i(b) = \exp^{x_i * b}$$

dan wordt de scoregelijkheid:

$$D \ln L = \sum(i) x_i * (1 - \sum(j \in R(t_i)) \frac{z_j(b)}{\sum(j \in R(t_i)) z_j(b)}) = 0$$

Paragraaf 4.8 Afsluiting.

Het in paragraaf 4.4 beschreven voorbeeld geeft aan, dat de gemeten overlevingscurve van een cohort niets zegt over de mate van heterogeniteit van een populatie. Om hierover meer te kunnen zeggen is het noodzakelijk om de overlevingscurven van subgroepen (zie onderhavige hoofdstuk) of expliciet risicovariabelen (zie volgende hoofdstuk) te beschouwen.

HOOFDSTUK 5 **Risico.**

Paragraaf 5.1 **Inleiding.**

In de hoofdstukken 1 t/m 3 zijn slechts algemene of doodsoorzaak-specifieke overlevingsfuncties en hun eigenschappen beschreven. De kans op sterfte (risico) werd geassocieerd met de (cumulatieve) mortaliteitsdruk. Deze mortaliteitsdruk werd als een onafhankelijke grootte verondersteld. In hoofdstuk 4 is daaraan het concept heterogeniteit toegevoegd. Heterogeniteit houdt in, dat de overlevingscurve van een geboorte-cohort en de overlevingsfunctie van een willekeurig individu uit dit cohort kunnen verschillen. Voor een heterogeen cohort kan verandering van de gemeten (collectieve) mortaliteitsdruk ook veroorzaakt worden door selectie in plaats van veranderingen in de mortaliteitsdrukken van de afzonderlijke individuen.

Heterogeniteit houdt in, dat de individuen van een cohort een verschillend risico bezitten. Het risico is in paragraaf 1.3 gedefinieerd als de cumulatieve mortaliteitsdruk. De individuen van een heterogeen cohort verschillen in hun mortaliteitsdruk. Het verschil tussen een benadering in termen van heterogeniteit en van risico is niet principieel. In het algemeen kan gezegd worden, dat bij heterogeniteit het (verschil in) risico impliciet uit de overlijdenstijdstippen bepaald wordt, en bij een risico-benadering de waarden van specifieke risicovariabelen expliciet gemeten worden. Dit onderscheid is echter niet fundamenteel en wordt vaak meer bepaald door conventies. Zo zijn de multiplicatief-risico benadering en de benadering van heterogeniteit met behulp van een mixing variable vrijwel identiek.

In dit hoofdstuk wordt de nadruk gelegd op het vergelijken van subpopulaties en het toetsen van hypothesen met betrekking tot (verhoogd) risico (zie bijvoorbeeld ook Cox ea, 1984). Meestal gebeurt dit door een studiegroep te vergelijken met een controle-groep of een gehele populatie. Hypothesen worden (in dit geval) geformuleerd in de vorm van veronderstellingen over de parameter-waarde(n) van een (geparametriseerde) overlevingsfunctie of mortaliteitsdruk van een studiegroep in relatie tot een controle-groep. Verschillende mogelijkheden om een (verhoogd) risico te beschrijven worden genoemd.

Paragraaf 5.2 **De accelerated-life beschrijving.**

De accelerated-life beschrijving houdt in, dat een persoon versneld veroudert (x : gegeven stochastisch overlijdenstijdstip, x' : met betrekking tot x accelerated-life overlijdenstijdstip, $c > 1$):

$$\Pr(x' \leq x) = \Pr(x \leq c \cdot x') = \Pr(x/c \leq x'), \text{ ofwel:}$$

$$x' = x/c$$

Een aantal andere grootheden kan eenvoudig in de overeenkomstige grootheden van de referentie-functie uitgedrukt worden:

$$S'(x) = S(c \cdot x)$$

$$h'(x) = f'(x) / S'(x) = c \cdot h(c \cdot x)$$

$$H'(x) = H(c \cdot x)$$

$$e'(x) = e(c \cdot x) / c$$

De laatste grootheid karakteriseert de accelerated life beschrijving het meest duidelijk: een dubbele afname van de levensverwachting. Het lineaire verband tussen de gegeven en accelerated-life overlijdenstijdstippen wordt een constant verschil voor de logaritmen ervan:

$$\ln x' = \ln x - c$$

Deze eigenschap biedt de mogelijkheid om de accelerated-life beschrijving grafisch te toetsen.

Paragraaf 5.3 De multiplicatief-risico beschrijving.

De multiplicatief-risico (Engels: proportional hazards) beschrijving houdt in, dat de mortaliteitsdruk ten opzichte van de referentie-mortaliteitsdruk een constante fractie groter is (x : gegeven stochastisch overlijdenstijdstip, x' : overlijdenstijdstip volgens multiplicatief-risico beschrijving, $c > 1$):

$$h'(x) = c \cdot h(x)$$

Een aantal van de andere grootheden, uitgedrukt in de overeenkomstige referentie-grootheden, wordt:

$$S'(x) = S(x)^c$$

$$f'(x) = c S(x)^{c-1} f(x)$$

$$e'(x) = \int_x^\infty S'(s) ds / S'(x) = \int_x^\infty S(x)^c ds / S(x)^c$$

De vergelijking voor de resterende levensverwachting geeft aan, dat de toename van de levensverwachting ten gevolge van een constante relatieve afname van de mortaliteitsdruk (ofwel: een multiplicatief risico kleiner dan 1) voor de Nederlandse bevolking gering zal zijn. Dit wordt veroorzaakt door de rechthoekige vorm van de Nederlandse overlevingscurve (zie ook paragraaf 8.8).

Het verband tussen de stochastische overlijdenstijdstippen x' en x is niet eenvoudig weer te geven. Daarentegen zijn aannemelijkheidsfuncties voor het toetsen van de beschrijving wel eenvoudig te formuleren. De aannemelijkheidsfunctie van de overlijdenstijdstippen x_i van een groep van n onderling onafhankelijke, identieke personen luidt onder de aanname van een multiplicatief risico c :

$$L(x|c) = \prod_{i \in N} e^{-H'(x_i)} \prod_{i \in M} h'(x_i)$$

$$l(x|c) = \#(M) \ln c + \sum_{i \in M} \ln h(x_i) - c \sum_{i \in N} H(x_i)$$

$$D l(x|c) = \frac{\#(M)}{c} - \sum_{i \in N} H(x_i) = 0$$

$$\hat{c} = \frac{\#(M)}{\sum_{i \in N} H(x_i)}$$

De grootheid $\#(M) / \sum_{i \in N} H(x_i)$ wordt het gestandaardiseerde sterftecijfer (SMR) genoemd.

Interessant is de relatie tussen de accelerated-life en de multiplicatief-risico beschrijving, ofwel: voor welke klasse(n) van verdelingen zijn beide gelijkwaardig? Om deze verdelingen te bepalen leiden we de volgende reeks gelijkheden af:

$$\text{Veronderstelling: } S(x)^c = S(d \cdot x)$$

$$\text{Pas een transformatie toe: } \ln c + \ln(-\ln S(x)) = \ln(-\ln S(d \cdot x))$$

$$\text{Substitueer: } f(y) = \ln(-\ln S(\exp(y))) : \ln c + f(y) = f(y + \ln d)$$

Aan deze laatste gelijkheid wordt voor alle y voldaan, indien $f(y)$ een lineaire functie is met richtingscoëfficiënt $\ln c / \ln d$. Terugsubstitutie en herdefiniëring van de constanten levert op:

$$S(x) = \exp(-a x^b)$$

ofwel: de klasse van Weibull-verdelingsfuncties is de enige, waarvoor de multiplicatief-risico en accelerated-life beschrijving in elkaar om te zetten zijn.

Paragraaf 5.4 De additief-risico beschrijving.

De additief-risico beschrijving geeft aan, dat tussen de mortaliteitsdruk van het referentie-overlijdenstijdstip en die van het te onderzoeken tijdstip een constant verschil bestaat (x : gegeven stochastisch overlijdenstijdstip, x' : overlijdenstijdstip volgens additief-risico beschrijving, $c > 0$):

$$h'(x) = c + h(x)$$

Hieruit volgt voor de overlevingsfunctie:

$$S'(x) = e^{cx} S(x)$$

Slechts voor exponentiële kansverdelingen kan deze additief-risico beschrijving omgezet worden in een accelerated-life of- multiplicatief-risico beschrijving. In dit geval kan de constante verandering van de mortaliteitsdruk geïnterpreteerd worden als de toevoeging c.q. verwijdering van een doodsoorzaak.

Paragraaf 5.3 Gegeneraliseerde lineaire modellen van risico.

Verondersteld wordt, dat het (verhoogd) risico van een individu (heterogeniteit) beschreven kan worden in de vorm van een multiplicatieve factor met betrekking tot de basic intensity (zie paragraaf 4.5 en 5.3). Het individuele risico wordt evenredig verondersteld met de basic intensity (proportional hazard):

$$h_{x|b}(t) = z(b) * h_c(t)$$

In deze paragraaf wordt een aantal gegeneraliseerde lineaire modellen van de factor $z(b)$ gepresenteerd. In paragraaf 4.5 werd de factor z als een niet-negatieve stochastische variabele met verwachting 1 beschreven. In deze paragraaf wordt deze factor niet-negatief verondersteld en afhankelijk van de waarde(n) van de zogenaamde risicovariabele(n). Hierbij wordt gebruik gemaakt van de volgende symbolen:

b	de (vectoriële) parameter
x	de waarden van de risicovariabele(n)
$z(x b)$	de multiplicatieve risicofactor als functie van x voor gegeven b

De gegeneraliseerde lineaire modellen kunnen worden beschreven als:

$$f(z(x|b)) = b'x$$

De functie f wordt de (Engels:) link function genoemd. Voorbeelden van gegeneraliseerde lineaire modellen zijn:

Multiplicatief risico (Engels: proportional hazards):

$$f(z) = \ln z$$

$$z(x|b) = \exp^{x'b} = \Pi(k) \exp^{xk+bk} \quad \text{met } 0 < z$$

Indien een risicovariabele z dichotoom wordt verondersteld, kan de grootheid $\exp(bk)$ geïnterpreteerd worden als het relatief risico van factor k .

Logistisch risico:

$$f(z) = \text{logit } z = \ln \frac{z}{1-z}$$

$$z(x|b) = \frac{\exp(b'x)}{1 + \exp(b'x)} \quad \text{met } 0 < z < 1$$

Complementair log-log:

$$f(z) = \ln(-\ln(1-z))$$

$$z(x|b) = 1 - \exp(-\exp(b'x)) \quad \text{met } 0 < z < 1$$

Probit:

$$f(z) = \Phi^{-1}(z) \quad (\Phi \text{ is de cumulatieve Normale verdeling})$$

$$z(x|b) = \Phi(b'x) \quad \text{met } 0 < z < 1$$

HOOFDSTUK 6 Concurrerende doodsoorzaken.

Paragraaf 6.1 Inleiding.

In dit hoofdstuk wordt de optredende sterfte gespecificeerd naar doodsoorzaak. Doodsoorzaken zijn concurrerend in die zin, dat slechts één oorzaak direct het overlijden tot gevolg heeft. Deze oorzaak wordt 'de' doodsoorzaak genoemd. Deze aanname heeft in de eerste plaats een theoretische achtergrond, maar sluit ook aan op het door het CBS gehanteerde onderscheid tussen primaire en secundaire doodsoorzaak. In geval van afwezigheid van de primaire doodsoorzaak zou de secundaire doodsoorzaak tot sterfte geleid hebben.

Het hoofdstuk is als volgt opgebouwd. Begonnen wordt met een algemene beschrijving van concurrerende doodsoorzaken in termen van een gemeenschappelijke kansverdelingsfunctie van overlijdenstijdstip en doodsoorzaak. Vervolgens wordt op basis van deze kansverdelingsfunctie de aannemelijkheidsfunctie geconstrueerd voor het algemene geval en in geval van type 1 en 2 censurering (zie bijvoorbeeld Mode, 1985).

Paragraaf 6.2 Beschrijving van het sterfteproces van een individu.

In deze paragraaf wordt het sterfteproces van een willekeurig individu beschreven in termen van het overlijdenstijdstip x en de doodsoorzaak G , beide stochastische grootheden. Hierbij wordt gebruik gemaakt van de tweedimensionale kansverdelingsfunctie van beide. De volgende nieuwe grootheden worden geïntroduceerd:

G	het stochastische optreden van de doodsoorzaak
$G_j(x)$	de gemeenschappelijke kansverdelingsfunctie van het overlijdenstijdstip x en de doodsoorzaak G ,
$G(x)$	de marginale kansverdelingsfunctie van het overlijdenstijdstip
$G(j)$	de marginale kansverdelingsfunctie van doodsoorzaak G
$h_j(x)$	de mortaliteitsdruk voor doodsoorzaak G_j
$S_j(x)$	de overlevingsfunctie voor doodsoorzaak G_j

De gemeenschappelijke kansverdelingsfunctie $G_j(x)$ vormt in eerste instantie het uitgangspunt. De overige grootheden kunnen daaruit afgeleid worden:

$$G_j(x) = \Pr(X \leq x, G=j)$$

$$G(x) = \Pr(X \leq x) = \sum \Pr(X \leq x, G=j) = \sum(j) G_j(x)$$

$$G(j) = \Pr(G=j) = \lim_{x \rightarrow 0} \Pr(X \leq x, G=j) = G_j(0)$$

$$S(x) = 1 - G(x)$$

$$h_j(x) = \lim_{h \rightarrow 0} \frac{\Pr(X \leq x+h, G=j | X > x)}{h} = \frac{d}{dx} G_j(x) / S(x)$$

$$h(x) = \lim_{h \rightarrow 0} \frac{\Pr(X \leq x+h | X > x)}{h} \\ = \lim_{h \rightarrow 0} \sum(j) \frac{\Pr(X \leq x+h, G=j | X > x)}{h} = \sum(j) h_j(x)$$

$$S_j(x) = \exp^{-H_j(x)}$$

$h_j(x)$ en $S_j(x)$ worden de bruto (Engels: crude) mortaliteitsdruk resp. overlevingsfunctie voor doodsoorzaak G_j genoemd. De overlevingsfunctie $S(x)$ kan geschreven worden als het produkt van de bruto overlevingsfuncties voor de verschillende doodsoorzaken:

$$S(x) = \exp^{-H(x)} = \exp^{-\sum(j) H_j(x)} = \prod(j) \exp^{-H_j(x)} \\ = \prod(j) S_j(x)$$

Tenslotte nog een opmerking over de wederzijdse afleidbaarheid van de mortaliteitsdruk en de overlevingsfractie. In hoofdstuk 1 is aangetoond, dat de (algemene) mortaliteitsdruk en de (algemene) overlevingsfunctie uit elkaar afgeleid kunnen worden. Een analoog resultaat geldt in het geval van specifieke doodsoorzaken, namelijk:

$$h_j(x) = \frac{d}{dx} G_j(x) / S(x) \text{ resp.}$$

$$G_j(x) = \int_0^x \exp^{-H(s)} h_j(s) ds \text{ met } H(s) = \int_0^s \sum(j) h_j(t) dt$$

Paragraaf 6.3 De algemene samenlijksheidsfunctie.

In deze en de volgende paragrafen komen de (realisaties van de) sterfteprocessen van een aantal individuen aan bod. Deze worden met de volgende grootheden beschreven:

- i index met betrekking tot de verschillende individuen
 N de verzameling van alle individuen,
 $M(j)$ de verzameling van individuen, die sterven aan
doodsoorzaak C_j , met aantal $m(j)$,
 t_i het overlijdenstijdstip van individu i ,

De doodsoorzaak-voorwaardelijke kansdichtheidsfunctie van het overlijdenstijdstip van een willekeurig individu i luidt:

$$\begin{aligned} \frac{d}{dx} \text{Pri}(x < x \mid C=j) &= - \frac{d}{dx} \text{Pri}(x < x, C=j) / \text{Pri}(C=j) \\ &= \frac{d}{dx} G_j(x) / C_j(j) = h_j(x) * S_j(x) / C_j(j) \end{aligned}$$

De overlevingsfunctie $S_j(x)$ kan herschreven worden tot:

$$\begin{aligned} S_j(x) &= \exp^{-H_j(x)} = \exp^{-\int_0^x h_j(s) ds} = \exp^{-\int_0^x \sum_j h_j(s) ds} \\ &= \exp^{-\sum_j \int_0^x h_j(s) ds} = \Pi(j) \exp^{-\int_0^x h_j(s) ds} \end{aligned}$$

De grootte $M = (M(j))$ is een discrete stochastische variabele met multinomiale verdelingsfunctie:

$$\text{Pr}(M) = \frac{N!}{m(j)!} \prod_{j \in M(j)} C_j(j)$$

Aangenomen wordt, dat de overlijdenstijdstippen van de verschillende individuen onderling onafhankelijk zijn. De aannemelijkheidsfunctie L van de realisaties van deze overlijdenstijdstippen wordt:

$$L = \frac{N!}{m(j)!} \prod_{j \in M(j)} h_j(t_i) \exp^{-H_j(t_i)}$$

Deze aannemelijkheidsfunctie kan herschreven worden tot:

$$L = \text{cst} * \Pi(j) \left(\prod_{i \in N} \exp^{-H_j(t_i)} * \prod_{i \in M(j)} h_j(t_i) \right)$$

Omdat de aannemelijkheidsfunctie geschreven kan worden als het produkt over de verschillende doodsoorzaken van een reeks gelijkvormige termen wordt deze multiplicatief separabel genoemd.

Paragraaf 6.4 De aannemelijkheidsfunctie onder type 1 censurering.

De algemene aannemelijkheidsfunctie onder type 1 censurering wordt:

$$L_1 = \text{cst} \prod_{j \in M(j)} h_j(t_i) \exp^{-H_j(t_i)} * \prod_{i \in N-M} \exp^{-H_j(c_i)}$$

Deze aannemelijkheidsfunctie kan worden herschreven tot:

$$L1 = \text{cst} * \prod(j) \left(\prod_{i \in N} \exp^{-H_{ij}(T_i)} * \prod_{i \in N(j)} h_{ij}(T_i) \right)$$

waarbij T_i (zie paragraaf 2.3) het minimum van de maximale waarnemingsduur en het overlijdenstijdstip vorat: $T_i = \min(t_i, c_i)$.

Bij deze aannemelijkheidsfunctie kunnen twee opmerkingen geplaatst worden:

- Uit de vorm van de formule blijkt weer de multiplicatieve separabiliteit,
- Deze functie komt na substitutie van T_i door t_i overeen met de aannemelijkheidsfunctie zonder censurering.

Paragraaf 6.5 De aannemelijkheidsfunctie onder type 2 censurering.

De algemene aannemelijkheidsfunctie onder type 2. censurering wordt:

$$L1 = \text{cst} \prod_{j, i \in N(j)} h_{ij}(t_i) \exp^{-H_i(t_i)} * \prod_{i \in N-M} \exp^{-H_i(c_i)}$$

Deze aannemelijkheidsfunctie kan worden herschreven tot:

$$L2 = \text{cst} * \prod_{j, i \in N(j)} h_{ij}(t_i) \exp^{-H_i(t_i)} * \prod_{i \in N-M} \exp^{-H_i(t(m))}$$

Hieruit blijkt weer, dat indien voor type 1 censurering de waarnemingsperiode van alle individuen $c_i - t(m)$ gekozen wordt de aannemelijkheidsfuncties $L1$ en $L2$ identiek zijn.

Paragraaf 6.6 De partiële aannemelijkheidsfunctie.

In paragraaf 4.3 is afgeleid, dat de algemene aannemelijkheidsfunctie van de gerealiseerde overlijdenstijdstippen en doodsoorzaken van een aantal onderling onafhankelijke individuen multiplicatief separabel is met betrekking tot de doodsoorzaken. Zonder beperking van de algemeenheid kan dus de produktterm voor een willekeurige doodsoorzaak C_j in plaats van de volledige aannemelijkheidsfunctie uitgewerkt worden. De index j met betrekking tot de doodsoorzaak wordt daarom in deze produktterm weggelaten. De (produktterm van de) aannemelijkheidsfunctie wordt hiermee:

$$L = \prod_{i \in N} \exp^{-H_i(t_i)} * \prod_{i \in N} h_i(t_i)$$

Deze aannemelijkheidsfunctie voor een willekeurige doodsoorzaak C_j komt overeen met de bekende algemene aannemelijkheidsfunctie, waaruit een partiële aannemelijkheidsfunctie afgeleid kan worden: ...

$$L = \prod_{i \in N} \exp^{-H_i(t_i)} * \prod_{i \in M} h_i(t_i) \\ = \prod_{i, j \in R(t_i)} \exp^{-\int_{t_{i-1}}^{t_i} h_j(s) ds} * \prod_{j \in K(t_i)} h_j(t_i) * \\ \prod_{i \in \bar{R}(t_i)} h_i(t_i)$$

De laatstgenoemde term wordt weer de partiële aannemelijkheidsfunctie genoemd.

Paragraaf 6.7 Subject-intervallen.

De partiële aannemelijkheidsfunctie ontstaat door de observatieperiode van een cohort op te delen naar de stochastische overlijdenstijdstippen. Daarnaast kan de observatieperiode ook deterministisch opgedeeld worden in zogenaamde subject-intervallen (Engels: subject-intervals, subject-years, person-years). Ieder subject-interval heeft betrekking op een specifiek individu. Deze kan tijdens het tijdsinterval sterven, overeenkomend met een stochastische breedte van het interval, of overleven, overeenkomend met de maximale deterministische breedte van het interval. Verondersteld wordt, dat de doodsoorzaak-specifieke mortaliteitsdrukken van ieder individu gedurende een tijdsinterval constant zijn (aanscherping van de proportional hazards aanname) (zie Hempenius e.a., 1981).

Deze beschrijving komt overeen met de praktische situatie, waarin van de individuen van een cohort op vaste tijdstippen een aantal risicovariabelen resp. de overlijdenstijdstippen vastgelegd wordt. Over het niveau van de risicovariabelen tussen de opeenvolgende tijdstippen is niets bekend.

Een aantal bekende symbolen krijgt nu een nieuwe betekenis in termen van subject-intervallen:

N	de verzameling van subject-intervallen,
M_j	de verzameling van subject-intervallen, waarin sterfte aan doodsoorzaak C_j plaatsvindt,
i	index met betrekking tot de verschillende subject-intervallen,
w_i	de breedte van subject-interval i ,

h_{ij} constante mortaliteitsdruk voor doodsoorzaak j
gedurende subject-interval i

De oorspronkelijke aannemelijkheidsfunctie kan nu worden herschreven tot:

$$L = \text{cst} * \prod_j \left(\prod_{i \in N} \exp^{-w_i h_{ij}} * \prod_{i \in M_j} h_{ij} \right)$$

Paragraaf 6.8 Een voorbeeld: het proportional hazards model.

In deze paragraaf wordt een voorbeeld gegeven van de werking van het principe van concurrerende doodsoorzaken. Dit voorbeeld betreft het zogenaamde proportional hazards model. Dit model volgt uit de aanname:

De relatieve mortaliteitsdruk voor een specifieke doodsoorzaak ten opzichte van de totale mortaliteitsdruk is voor alle leeftijdjaren constant,

ofwel:

$$\text{Voor alle } j \in J, x \geq 0: h_j(x) = c_j * h(x)$$

Er geldt in dit geval:

$$G(j) = \int_0^\infty S(x) h_j(x) dx = c_j * \int_0^\infty S(x) h(x) dx = c_j$$

$$G_j(x) = \Pr(X \leq x, G=j) = \int_0^x S(s) h_j(s) ds$$

$$= c_j * \int_0^x S(s) h(s) ds = c_j * F(x) = \Pr(X \leq x) * \Pr(G=j)$$

ofwel: het tijdstip van overlijden en de doodsoorzaak zijn onderling onafhankelijk.

Omgekeerd volgt de proportional hazards aanname uit de veronderstelling, dat overlijdenstijdstip en doodsoorzaak onderlinge onafhankelijk zijn:

$$\begin{aligned} h_j(x) &= \lim_{h \downarrow 0} \frac{P(x < X \leq x+h, G=j \mid X > x)}{h} \\ &= \lim_{h \downarrow 0} \frac{P(x < X \leq x+h, G=j)}{h} / P(X > x) \\ &= \lim_{h \downarrow 0} \frac{P(x < X \leq x+h)}{h} * P(G=j) / P(X > x) = c_j * h(x) \end{aligned}$$

Deze gelijkwaardigheid van de proportional hazards aanname en de onderlinge onafhankelijkheid van doodsoorzaak en overlijdenstijdstip speelt een rol bij het onderscheid tussen concurrerende en vervangende doodsoorzaken (zie paragraaf 7.9).

Het proportional hazards model wordt toegepast in het actuariële model van concurrerende doodsoorzaken. Hierbij wordt uitgegaan van een raamwerk (Engels: grid) in de vorm van een reeks van discrete leeftijdsjaren. Gegeven dit raamwerk wordt voor ieder interval $[x_i, x_{i+1})$ tussen twee leeftijdsjaren (Engels: grid points) x_i resp. x_{i+1} het continue sterfteproces beschreven in de vorm van een proportional hazards model. In hoofdstuk 8 wordt deze versie van het proportional hazards model uitgewerkt.

HOOFDSTUK 7 Virtuele overlijdenstijdstippen.

Paragraaf 7.1 Inleiding.

In hoofdstuk 6 is het concept concurrerende doodsoorzaken geïntroduceerd. Hierbij wordt het sterfteproces van een individu met behulp van een tweewaardige stochastische variabele beschreven: (x, G) . De variabele x beschrijft het overlijdenstijdstip, de variabele G de doodsoorzaak. Op basis hiervan kan de aannemelijkheidsfunctie van de realisaties van overlijdenstijdstip en doodsoorzaak beschreven worden. Deze aannemelijkheidsfunctie is multiplicatief separabel met betrekking tot de doodsoorzaken.

In dit hoofdstuk wordt het concept virtuele (Engels: latent) overlijdenstijdstip geïntroduceerd. Het virtuele overlijdenstijdstip wordt toegepast om een meer abstracte beschrijving van het stochastisch gedrag van de genoemde tweewaardige stochastische variabele (x, G) te geven. Voor ieder individu wordt voor iedere onderscheiden doodsoorzaak een stochastisch tijdstip gegeven, te interpreteren als het tijdstip van overlijden aan de betreffende oorzaak. Het minimum van deze overlijdenstijdstippen vormt het tijdstip van overlijden. De oorzaak, waarvoor het overlijdenstijdstip en het genoemde minimum overeenkomen, wordt de doodsoorzaak genoemd. De overige overlijdenstijdstippen zijn gecensureerd en kunnen niet vaargenomen worden.

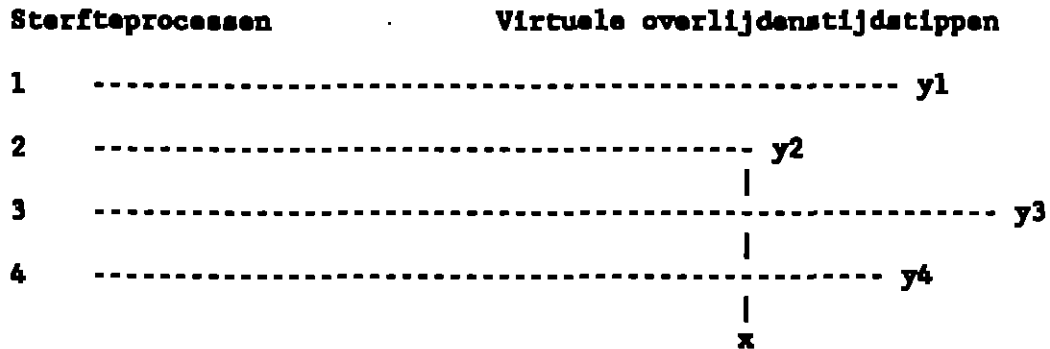
Paragraaf 7.2 Introductie van het concept.

De volgende (nieuwe) grootheden worden nu (voor een willekeurig individu) geïntroduceerd:

- $y = (y_j)$ de stochastische overlijdenstijdstippen met betrekking tot de (mogelijke) doodsoorzaken G_j ,
- $\tilde{S}(y)$ de gemeenschappelijke verdelingsfunctie van de stochastische variabele y ,
- $S(x)$ de verdelingsfunctie van het werkelijke overlijdenstijdstip x , ofwel: de leeftijd-afhankelijke overlevingskans,
- $\tilde{S}_j(y_j)$ de j -de marginale overlevingsfunctie
 $= \tilde{S}(y_j * e_j)$ (e_j is de j -de eenheidsvector)

$\tilde{S}_j(x)$ wordt de netto (Engels: net) overlevingsfunctie genoemd. Vergelijk hiermee de bruto overlevingsfunctie $S_j(x)$ (zie paragraaf 6.2).

In het volgende schema wordt een voorbeeld van het principe van de virtuele overlijdenstijdstippen gegeven.



In dit voorbeeld vormt de tweede oorzaak de doodsoorzaak en is het overlijdenstijdstip x gelijk aan het gerealiseerde overlijdenstijdstip y_2 . De overige tijdstippen y_1 , y_3 en y_4 zijn gecensureerd.

Zoals gezegd vormt het stelsel stochastische variabelen $y = (y_j)$ de basis van het sterfteproces. De overige genoemde grootheden kunnen als volgt uit dit stelsel afgeleid worden:

$$\begin{aligned}
 x &= \min(y_j \mid j \in J) \\
 G-j &\text{ desda } x = y_j, \text{ ofwel } G = \arg(\min(y_j \mid j \in J)) \\
 G_j(x) &= \Pr(x \leq x, G-j) = \Pr(y_j \leq x, y_j \leq y_k \mid k \neq j) \\
 \tilde{S}(y) &= \Pr(y_j > y_j \mid j \in J) \\
 S(x) &= \Pr(x > x) = \Pr(\min(y_j \mid j \in J) > x) \\
 &= \Pr(y_j > x \mid j \in J) = \tilde{S}(x, \underline{1}) \quad (\underline{1} \text{ is de éénvector})
 \end{aligned}$$

Indien doodsoorzaak G_j de doodsoorzaak is en y_j het minimum van het stelsel $y = (y_j)$, wordt j ook wel de minimaliserende index genoemd. De beschrijving van de overlevingsfunctie in termen van de virtuele overlijdenstijdstippen geeft aan, dat het sterfteproces van een individu beschouwd kan worden als een reeks afzonderlijke sterfteprocessen voor de verschillende doodsoorzaken, waarbij de tijden gesynchroniseerd verlopen.

Paragraaf 7.3 Uitwerking van het concept.

In deze paragraaf wordt het in paragraaf 7.2 geïntroduceerde principe van de virtuele overlijdenstijdstippen voor een willekeurig individu uitgewerkt.

De meerdimensionale mortaliteitsdruk:

$R(y)$ de risicofunctie (cumulatieve mortaliteitsdruk)

$n(y)$ de risicogradient (mortaliteitsdruk)

$R(y)$ - - $\ln \tilde{S}(y)$

$n(y)$ - grad $R(y)$

De totale mortaliteitsdruk:

$R(y)$ - $\int_0^y n(s) ds$

$\tilde{S}(y)$ - $\exp^{-R(y)} = \exp^{-\int_0^y n(s) ds}$

$h_j(x)$ - $-\frac{d}{dy_j} \tilde{S}(x^*1) / \tilde{S}(x^*1) = -\frac{d}{dy_j} \tilde{S}(x^*1) = \frac{d}{dy_j} R(x^*1)$
 - $n_j(x^*1)$

$h(x)$ - $\sum(j) h_j(x) = \sum(j) n_j(x^*1) = \text{inprodukt}(n(x^*1), 1)$

$h_j(x)$ wordt de bruto mortaliteitsdruk (Engels: crude hazard rate) voor doodsoorzaak C_j genoemd (zie ook paragraaf 6.2). Naast de bruto mortaliteitsdruk wordt ook een netto (Engels: net) mortaliteitsdruk $k_j(x)$ onderscheiden op basis van de netto overlevingsfunctie $\tilde{S}_j(x)$:

$k_j(x) = -\frac{d}{dx} \ln \tilde{S}_j(x)$

$\tilde{S}_j(x) = \exp^{-\int_0^x k_j(s) ds}$

De bruto mortaliteitsdruk $h_j(x)$ beschrijft de (voorwaardelijke) kans om onder invloed van alle doodsoorzaken aan doodsoorzaak C_j te overlijden op tijdstip x , de netto mortaliteitsdruk $k_j(x)$ de kans om onder invloed van slechts doodsoorzaak C_j te overlijden. In het algemeen zijn beide kansen verschillend (zie ook paragraaf 7.5), omdat verondersteld mag worden, dat de processen, die leiden tot sterfte aan de verschillende doodsoorzaken, elkaar beïnvloeden.

Paragraaf 7.4 Sterke onderlinge onafhankelijkheid.

In de voorafgaande beschrijving werd ervan uitgegaan, dat de virtuele overlijdenstijdstippen in principe onderling afhankelijk (kunnen) zijn. In deze paragraaf worden de gevolgen onderzocht van de vereenvoudigende aanname, dat de overlijdenstijdstippen (sterk) onderling onafhankelijk zijn. Er geldt:

$\mathbf{y} = (y_j)$ vormt een stelsel (sterk) onderling onafhankelijke stochastische variabelen

desda

$\tilde{S}(\mathbf{y}) = \prod(j) \tilde{S}_j(y_j)$, de gemeenschappelijke overlevingsfunctie vormt het produkt van de marginale overlevingsfuncties voor alle tijdstippen y

en dus na integreren: $S_j(x) = \tilde{S}_j(x)$. De bruto en netto mortaliteitsdruk resp. overlevingsfunctie zijn voor alle onderscheiden doodsoorzaken identiek. Ofwel: de invloed van het aan een specifieke doodsoorzaak ten grondslag liggende proces heeft slechts invloed op de sterfte aan de betreffende doodsoorzaak en niet op de overige sterfte.

Paragraaf 7.5 Zwakke onderlinge onafhankelijkheid.

De stochastische virtuele overlijdenstijdstippen worden (sterk) onderling onafhankelijk genoemd, indien de gemeenschappelijke overlevingsfunctie het produkt vormt van de marginale overlevingsfuncties. Deze voorwaarde moet gelden voor alle mogelijke gemeenschappelijke (J -dimensionale) tijdstippen. In werkelijkheid is echter slechts sprake van één tijdsverloop; de J verschillende tijdsvariabelen lopen synchroon. De aanname van zwakke onderlinge onafhankelijkheid heeft op het laatste geval betrekking.

$\mathbf{y} = (y_j)$ vormt een stelsel zwak onderling onafhankelijke stochastische variabelen

desda

$S(x) = \tilde{S}(x+1) = \prod(j) \tilde{S}_j(x)$, de gemeenschappelijke overlevingsfunctie vormt het produkt van de marginale overlevingsfuncties voor de synchroon verlopende tijd x

Uit de definitie blijkt, dat een sterk onderling onafhankelijk stelsel eveneens zwak onderling onafhankelijk is, echter niet omgekeerd. De aanname van zwakke onderlinge onafhankelijkheid van de overlijdenstijdstippen is gelijkwaardig aan de aanname van gelijkheid van de bruto en netto mortaliteitsdruk:

$$h_j(x) = - \frac{d}{dy_j} \ln \tilde{S}(x+1) = - \frac{d}{dx} \ln \tilde{S}_j(x) = k_j(x)$$

resp.

$$\begin{aligned} \tilde{S}(x+1) &= S(x) = \Pi(j) S_j(x) = \Pi(j) \exp^{-H_j(x)} = \Pi(j) \exp^{-K_j(x)} \\ &= \Pi(j) \tilde{S}_j(x) \end{aligned}$$

In het algemeen zullen $h_j(x)$ en $k_j(x)$ echter verschillend zijn, omdat verwacht mag worden, dat de wijze van overleving in het verleden, onder invloed van alle doodsoorzaken óf slechts onder invloed van oorzaak C_j , van belang is.

Paragraaf 7.6 Niet-identificeerbaarheid.

In geval van (zwak en daarmee ook sterk) onderling onafhankelijke virtuele overlijdenstijdstippen kan de invloed van de j -de doodsoorzaak geïdentificeerd worden met de j -de marginale overlevingsfunctie. De marginale overlevingsfuncties worden vastgelegd door de doodsoorzaak-specifieke mortaliteitsdrukken:

$$- \frac{d}{dx} \ln \tilde{S}_j(x) = k_j(x) = h_j(x)$$

$$\tilde{S}_j(x) = \exp^{-\int_0^x h_j(s) ds}$$

Twee stelsels van virtuele overlijdenstijdstippen $\underline{y} = (y_j)$ en $\underline{x} = (x_j)$ worden equivalent genoemd, indien de hiervan afgeleide tweewaardige stochastische variabelen $(\underline{x}, G)$, ofwel de bijbehorende tweedimensionale verdelingsfuncties $G_j(x)$ identiek zijn.

Omgekeerd bestaat echter een keuzemogelijkheid bij het definiëren van een stelsel virtuele overlijdenstijdstippen vanuit een tweedimensionale verdelingsfunctie van overlijdenstijdstip en doodsoorzaak. De mogelijk te definiëren stelsels zijn equivalent. Indien echter verondersteld wordt, dat de virtuele overlijdenstijdstippen sterk onderling onafhankelijk zijn, wordt de meerdimensionale overlevingsfunctie eenduidig bepaald:

$$\tilde{S}(\underline{x}) = \Pi(j) \tilde{S}_j(x_j) = \Pi(j) \exp^{-\int_0^{x_j} h_j(s) ds}$$

Deze vrijheid van keuze van de meerdimensionale overlevingsfunctie wordt aangeduid met de term non-identificeerbaarheid van virtuele overlijdenstijdstippen. Non-identificeerbaarheid vormt een duidelijke beperking van toepassing van het principe van de virtuele overlijdenstijdstippen. In geval van (sterke) onderlinge onafhankelijkheid kan de invloed van de j -de doodsoorzaak geïdentificeerd worden met de j -de marginale overlevingsfunctie. Worden de doodsoorzaak-specifieke mortaliteitsdrukken van gemeenschappelijke risicovariabelen afhankelijk verondersteld, dan

zijn de overlijdenstijdstippen in het algemeen niet meer onderling onafhankelijk.

Paragraaf 7.7 Een voorbeeld van een keuzemogelijkheid.

De multiplicatieve separabiliteit van de samentelikhedsfunctie met betrekking tot de verschillende doodsoorzaken wil niet zeggen, dat deze doodsoorzaken onderling onafhankelijk zijn. Anders gezegd: de doodsoorzaak-specifieke overlijdenstijdstippen van de individuen van een cohort kunnen beschreven worden door onafhankelijke en door afhankelijke mortaliteitsdrukken. Het volgende voorbeeld maakt dit duidelijk.

We beschouwen een cohort individuen, die bij de geboorte een reeks stochastische doodsoorzaak-specifieke overlijdenstijdstippen toegewezen krijgen. De verdelingsfuncties hiervan worden beschreven door de doodsoorzaak-specifieke overlevingsfuncties. De realisaties van deze tijdstippen bepalen de gemiddelde doodsoorzaak-specifieke overlevingscurves van een cohort, die gemeten kunnen worden. Omgekeerd kunnen uit gemeten overlevingscurves (de) achterliggende overlevingsfuncties bepaald worden. Analooog aan het geval van heterogeniteit van een cohort (paragraaf 4.4) bezit de onderzoeker een keuzevrijheid. Bijvoorbeeld: men kan veronderstellen, dat de toegewezen overlijdenstijdstippen onderling onafhankelijk, of dat deze onderling afhankelijk zijn. Het voorbeeld beschrijft deze keuzemogelijkheid.

Twee doodsoorzaken worden onderscheiden, beide met gemeten gemiddelde mortaliteitsdruk:

$$h_1(t) = h_2(t) = \frac{1}{1+2t}$$

Deze mortaliteitsdrukken kunnen het gevolg zijn van:

Identieke stochastische mortaliteitsdrukken

Ieder individu krijgt een constante mortaliteitsdruk r toegewezen, die voor beide doodsoorzaken geldt. Deze mortaliteitsdruk is negatief exponentieel verdeeld met parameter 1:

$$h_1(t) = h_2(t) = h(t) / 2 = - \frac{d}{dt} \ln E \left(\exp^{- \int_0^t 2 h_1(s) ds} \right) / 2$$

$$= \frac{1}{1+2t}$$

Onderling onafhankelijke stochastische mortaliteitsdrukken

Ieder individu krijgt de doodsoorzaak-specifieke mortaliteitsdrukken toegewezen:

$$h1_{z1}(t) = \frac{z1}{1+z1t}, \quad h2_{z2}(t) = \frac{z2}{1+z2t}$$

De parameters $z1$ en $z2$ zijn onderling onafhankelijk negatief exponentieel verdeeld met parameter 1:

$$h1(t) = h2(t) = h(t) / 2$$

$$= - \frac{d}{dt} \ln E \left(\exp^{- \int_0^t h1_{z1}(s) + h2_{z2}(s) ds} \right) / 2$$

$$= - \frac{d}{dt} \ln E \left(\exp^{- \int_0^t h1_{z1}(s) ds} \right)^2 / 2 = \frac{1}{1+2t}$$

Beide beschrijvingen leiden tot de gegeven gemiddelde mortaliteitsdrukken.

Paragraaf 7.8 De algemene aannemelijkheidsfunctie.

In de voorafgaande paragrafen werd het principe van de virtuele overlijdenstijdstippen beschreven voor een willekeurig individu. In deze paragraaf wordt de algemene aannemelijkheidsfunctie geschreven van de realisaties van de overlijdenstijdstippen in termen van het genoemde principe.

Hierbij wordt, naast de reeds in paragraaf 4.3 geïntroduceerde grootheden, van de volgende nieuwe grootheden gebruik gemaakt:

y_{ij} het overlijdenstijdstip van individu i voor doodsoorzaak C_j

De doodsoorzaak-voorwaardelijke kansdichtheidsfunctie van het overlijdenstijdstip voor een willekeurig individu i luidt:

$$\begin{aligned} \frac{d}{dx} \text{Pri}(x < x \mid G_j) &= - \frac{d}{dx} \text{Pri}(x > x \mid G_j) / \text{Pri}(G_j) \\ &= - \frac{d}{dx} \int_{y_j-x}^{\infty} \int_{y_k-y_j}^{\infty} \bar{s}_i(y) dy / C_i(j) \\ &= \int_{y_k-x}^{\infty} \bar{s}_i(x \mid y_j-x) \prod_j dy_k / C_i(j) \end{aligned}$$

De gemeenschappelijke kansdichtheidsfunctie van de onderling onafhankelijke overlijdenstijdstippen van alle individuen luidt:

$$\prod_{i \in H(j)} \int_{y_k-x}^{\infty} \bar{s}_i(x \mid y_j-t_i) \prod_j dy_k / C_i(j)^{n(j)}$$

ofwel, na invulling van de multinomiale verdeling van de overlijdensaantallen wordt de aannemelijkheidsfunctie:

$$L = \sum_{j \in M(j)} \frac{m!}{k!} \int_{y_k=x}^{\infty} \tilde{s}_i(x | y) - t_i \int_{k!}^{\infty} dy$$

In geval van onderlinge onafhankelijkheid van de virtuele overlijdenstijdstippen kan de algemene aannemelijkheidsfunctie herschreven worden tot:

$$L = \text{const} * \Pi(j) \left(\prod_{i \in M(j)} \tilde{s}_{ij}(t_i) * S_i(t_i) / \tilde{S}_{ij}(t_i) \right)$$

De laatste aannemelijkheidsfunctie vormt het produkt van een reeks gelijkvormige aannemelijkheidsfuncties voor de afzonderlijke doodsoorzaken. Vanwege deze eigenschap wordt de aannemelijkheidsfunctie multiplicatief separabel met betrekking tot de doodsoorzaken genoemd. Deze separabiliteit is echter een andere dan de separabiliteit, beschreven in paragraaf 4.3.

Paragraaf 7.9 Vervangende en concurrerende doodsoorzaken.

Onderscheid wordt gemaakt tussen vervangende en concurrerende doodsoorzaken. Dit onderscheid wordt vaak in verband gebracht met de (veronderstelde) gevolgen van eliminatie van een doodsoorzaak, getuige o.a.:

'The problem is whether one can assume that the population "saved" by the elimination of a cause of death can simply be returned to the pool of survivors and therefore runs the same risks of dying from the other causes of death.' (Gunning-Schepers, 1988)

In het licht hiervan worden doodsoorzaken vervangend genoemd, indien verandering van de mortaliteitsdruk voor een willekeurige doodsoorzaak geen invloed heeft op de mortaliteitsdrukken voor de overige doodsoorzaken. Doodsoorzaken worden concurrerend genoemd, indien ze niet vervangend zijn. Bijvoorbeeld: de veronderstelling, dat twee doodsoorzaken van een gemeenschappelijke risicofactor afhankelijk zijn, maakt de doodsoorzaken concurrerend.

Het onderscheid tussen vervangende en concurrerende doodsoorzaken wordt niet in termen van virtuele overlijdenstijdstippen geformuleerd en staat daarom los van het onderscheid tussen (sterk of zwak) onderling on- en afhankelijke virtuele overlijdenstijdstippen. Ook bestaat geen direct verband met de proportional hazards aanname, die een constante verhouding tussen de doodsoorzaak-specifieke mortaliteitsdrukken oplegt. Het staat de onderzoeker vrij om eliminatie van een doodsoorzaak gepaard te laten gaan met een (éénmalige) verandering in deze verhoudingen. Tenslotte bestaat ook geen verband tussen (sterk) onderling onafhankelijke overlijdenstijdstippen en de proportional hazards aanname.

HOOFDSTUK 8 Het actuariële model.

Paragraaf 8.1 Inleiding.

Het actuariële model vormt een beschrijving van concurrerende doodsoorzaken op basis van de proportional hazards aanname (zie bijvoorbeeld Chiang, 1968). Het model geeft een beschrijving in cohort-vorm: een groep van individuen zonder instroom wordt in de loop van de waarnemingstijd gevolgd. Een verandering van de waarnemingstijd komt aldus overeen met een verandering van leeftijd.

Hierbij wordt uitgegaan van een raamwerk (Engels: grid) in de vorm van een reeks van discrete leeftijdsjaren x_i (Engels: grid points). Voor ieder leeftijdsjaar x_i zijn de overlevingsaantallen en de aantallen overledenen aan de specifieke doodsoorzaken gedurende het voorafgaande leeftijdsinterval gegeven. Gegeven dit raamwerk wordt voor ieder interval (x_i, x_{i+1}) tussen twee leeftijdsjaren x_i en x_{i+1} het continue sterfteproces beschreven. Hierbij worden de volgende aannames gehanteerd:

- Per doodsoorzaak geldt voor alle overlevenden een zelfde mortaliteitsdruk, waarbij tussen de overlevenden geen afhankelijkheid optreedt. Ofwel: de sterfteprocessen zijn onderling onafhankelijk en identiek verdeeld.
- De sterfteprocessen zijn (op intervalniveau) geheugenloos. Dat wil zeggen: de sterfteaantallen gedurende een tijdsinterval zijn slechts afhankelijk van het aantal nog levenden aan het begin ervan en niet van de daaraan voorafgaande aantallen.

De proportional hazards aanname heeft betrekking op het sterfteproces over een willekeurig interval tussen twee leeftijdsjaren:

De relatieve mortaliteitsdruk voor een specifieke doodsoorzaak ten opzichte van de totale mortaliteitsdruk is over een interval constant.

Paragraaf 8.2 Een overzicht van de gehanteerde grootheden.

De volgende (nieuwe) grootheden met bijbehorende notatie worden gehanteerd:

C_j	de verschillende doodsoorzaken, $j=1..J$
x_i	de grens van een leeftijdsinterval
t	tijdvariabele, $x_i \leq t < x_{i+1}$

d_i	het aantal individuen, dat over het interval $[x_i, x_{i+1})$ sterft,
d_{ij}	als d_i met toegevoegd: aan doodsoorzaak C_j
p_i, q_i	de kans, dat een op x_i nog levend individu op x_{i+1} nog leeft resp. voor x_{i+1} sterft,
q_{ij}	als q_i met toegevoegd: aan doodsoorzaak C_j .
p_{ij}, q_{ij}	als p_i resp. q_i met toegevoegd: in aanwezigheid van slechts doodsoorzaak C_j
$p_{i,j}, q_{i,j}$	als p_i resp. q_i met toegevoegd: in afwezigheid van doodsoorzaak C_j
$h_j(s)$	de mortaliteitsdruk voor doodsoorzaak C_j
$h(s)$	de totale mortaliteitsdruk, $= \sum(j) h_j(s)$

De genoemde proportional hazards aanname kan nu geformuleerd worden als:

$$h_j(s) / h(s) = c_{ij} \quad \text{voor } x_i \leq s < x_{i+1}, j=1..J$$

Paragraaf 8.3 Beschrijving van de verschillende overlevings- en sterftekansen.

De verschillende overlevings en sterftekansen worden:

$$\begin{aligned}
 p_i &= \exp^{-\int_{x_i}^{x_{i+1}} h(s) ds} = \exp^{-H(x_i) - H(x_{i+1})} \\
 Q_{ij} &= \int_{x_i}^{x_{i+1}} \exp^{-\int_{x_i}^s h(t) dt} * h_j(s) ds = c_{ij} * q_i \\
 q_i &= 1 - p_i = \Pi(j) Q_{ij} \\
 p_{ij} &= \exp^{-\int_{x_i}^{x_{i+1}} h_j(s) ds} = \left(\exp^{-\int_{x_i}^{x_{i+1}} h(s) ds} \right)^{c_{ij}} \\
 &= p_i^{c_{ij}} \\
 q_{ij} &= 1 - p_{ij} \\
 &= 1 - (1 - q_i)^{c_{ij}} \quad (\text{onafhankelijke sterftekans}) \\
 &= c_{ij} q_i \quad (\text{afhankelijke sterftekans}) \\
 p_{i,j} &= p_i^{1-c_{ij}}
 \end{aligned}$$

$$\begin{aligned}
 q_{i,j} &= 1 - p_{i,j} \\
 &= 1 - (1 - q_i)^{1-c_{ij}} && \text{(onafhankelijke sterftekans)} \\
 &= (1 - c_{ij}) q_i && \text{(afhankelijke sterftekans)}
 \end{aligned}$$

- .. Bij zogenaamde afhankelijke sterftekansen wordt ervan uitgegaan, dat de doodsoorzaak-specifieke sterftekans gelijk is aan de waargenomen sterftefractie. De overige sterfte echter concurreert met de betreffende specifieke sterfte en leidt tot een vorm van censurering (zie paragraaf 2.5 e.v.). Een aantal van de overlijdenstijdstippen voor de betreffende doodsoorzaak wordt gecensureerd door de overige sterfte. De onafhankelijke sterftekans houdt hiermee rekening. De onafhankelijke sterftekans q_{ij} is groter dan de afhankelijke, omdat een deel van de overige sterfte meegerekend wordt. Ook de onafhankelijke sterftekans $q_{i,j}$ is groter dan de afhankelijke, omdat een deel van de personen, die anders zouden zijn overleden aan doodsoorzaak C_j , in geval van eliminatie van deze doodsoorzaak in dezelfde mate overlijden aan de andere doodsoorzaken als de andere overlevenden.
- .. In wiskundige termen: de afhankelijke kans vormt de eerste orde benadering van de onafhankelijke kans. Dit leidt ertoe, dat voor kleine sterftefracties q_i en proporties c_{ij} , in het algemeen optredend bij lage leeftijden resp. weinig voorkomende doodsoorzaken, beide kansen slechts weinig verschillen.

Nota bene: dit onderscheid tussen de onafhankelijke en afhankelijke sterftekans verschilt fundamenteel van het onderscheid tussen vervangende en concurrerende doodsoorzaken (zie paragraaf 7.9).

Tussen de verschillende sterftekansen bestaan de volgende ongelijkheids relaties:

$$Q_{ij} < q_{ij}$$

bewijs: voor $c_{ij}=0$ en $c_{ij}=1$ geldt: $Q_{ij} = q_{ij}$, Q_{ij} is lineair en q_{ij} concaaf in c_{ij} ,

interpretatie: zonder concurrentie is specifieke sterfte groter,

$$\text{als } Q_{ij} > Q_{ik} \text{ dan } q_{ij} > q_{ik}$$

bewijs: als $Q_{ij} > Q_{ik}$ dan $c_{ij} > c_{ik}$ enz,

interpretatie: een grotere mortaliteitsdruk uit zich in geval van aan- en afwezigheid van concurrentie,

Paragraaf 8.4 Het schatten van de kansen uit de sterfte-aantallen.

In de voorgaande paragrafen is een aantal overlevings- en sterftekansen bepaald als functie van de algemene mortaliteitsdruk- en de proportional hazards verhoudingen. Deze kansen kunnen geschat worden met behulp van de gerealiseerde sterfteaantallen. Hierbij wordt gebruik gemaakt van aannemelijkheidsfuncties.

Van een reeks waargenomen sterfteaantallen wordt de aannemelijkheidsfunctie opgesteld met als parameters de interval-specifieke overlevings- en sterftekansen. Verondersteld wordt, dat de individuen elkaar niet beïnvloeden en een identiek gedrag vertonen. Door maximalisatie van deze aannemelijkheidsfunctie kunnen de meest aannemelijke schatters van de overlevings- en sterftekansen afgeleid worden. De momenten van deze schatters worden bepaald aan de hand van de kansgenererende functie van de aannemelijkheidsfunctie.

Paragraaf 8.5 De kansverdelingen van de optredende sterfte.

De als eerste te bepalen kansverdeling is die van de optredende sterfte gedurende een leeftijdsinterval $[x_i, x_{i+1})$, gegeven het aantal overlevenden l_i aan het begin x_i van het interval. Omdat de overlijdenstijdstippen van de verschillende individuen onderling onafhankelijk en identiek verdeeld verondersteld worden, is deze voorwaardelijke kansverdeling multinomiaal:

$$\begin{aligned} \Pr((d_{ij}), l_{i+1} | l_i) \\ = \sum_{\{j\}} \frac{l_i!}{d_{i1}! \cdots d_{ij}! \cdots d_{iJ}!} \Pi(j) Q_{ij}^{d_{ij}} * p_i^{l_{i+1}} \end{aligned}$$

De bijbehorende voorwaardelijke kansgenererende functie vormt het produkt over de verschillende individuen van identieke individuele kansgenererende functies:

$$\begin{aligned} g((d_{ij}), l_{i+1} | l_i) &= (s_{ij}), s_{i+1} \\ &= E(\Pi(j) s_{ij}^{d_{ij}} * s_{i+1}^{l_{i+1}} | l_i) \\ &= (\Pi(j) Q_{ij} * s_{ij} + p_i * s_i + 1)^{l_i} \end{aligned}$$

De volgende te bepalen kansverdeling is de gemeenschappelijke kansverdeling van de optredende sterfteaantallen gedurende een opeenvolgende reeks van n leeftijdsintervallen, ofwel: gedurende een mensenleven. Deze kansverdeling is eveneens voorwaardelijk in die zin, dat de sterfteaantallen nu afhangen van het aantal individuen aan het begin van de reeks leeftijdsintervallen, ofwel: bij de geboorte. Deze verdeling luidt:

$$\text{Pr}((d_{ij}), l_{i+1} | 10)$$

$$= \Pi(i=0..n) \sum_{j=0}^{l_i} d_{ij}^{l_i} \cdot \Pi(j) Q_{ij}^{d_{ij}} \cdot p_i^{l_{i+1}}$$

Deze formule kan met behulp van volledige inductie naar het aantal opeenvolgende intervallen bewezen worden, waarbij gebruik wordt gemaakt van de geheugenloosheid van het optreden van de sterfte tijdens de opeenvolgende intervallen.

De bijbehorende gemeenschappelijke kansgenererende functie luidt:

$$G((d_{ij}), l_{i+1} | 10)((s_{ij}), s_{i+1})$$

$$= E(\Pi(i) \Pi(j) s_{ij}^{d_{ij}} \cdot s_{i+1}^{l_{i+1}} | 10)$$

$$= \left| \begin{array}{l} \sum_j Q_{0j} \cdot s_{0j} + \\ \sum_{i=1..n} (\sum_{j=0}^{l_{i-1}} p_{i-1} \cdot s_{i+1} \cdot \Pi_j Q_{i2j} \cdot s_{i2j}) + \\ \sum_{i=0..n} p_i \cdot s_{i+1} \end{array} \right|^{10}$$

Paragraaf 3.6 De schatters van de kansen.

De meest aannemelijke schatters van de verschillende overlevings- en sterftekansen worden afgeleid uit de aannemelijkheidsfunctie:

$$\Pi(i) \sum_{j=0}^{l_i} d_{ij}^{l_i} \cdot p_i^{l_{i+1}} \cdot \Pi(j) Q_{ij}^{d_{ij}} \quad \text{ofwel:}$$

$$\text{est} \cdot \Pi(i) p_i^{l_{i+1}} \cdot \Pi(j) Q_{ij}^{d_{ij}}$$

en daarvan de logarithme (met weglating van de constante):

$$\Sigma(i) l_{i+1} \cdot \ln p_i + \Sigma(i,j) d_{ij} \cdot \ln Q_{ij}$$

Van deze functie wordt het evenwichtspunt bepaald. Vanwege het (strikt) concave karakter van de (logarithme van de) aannemelijkheidsfunctie vormt dit evenwicht tevens het globale maximum. Hieruit volgen als meest aannemelijke schatters:

$$\hat{p}_i = l_{i+1} / l_i$$

$$\hat{Q}_{ij} = d_{ij} / l_i$$

$$\hat{p}_{ij} = \hat{p}_i \cdot d_{ij} / d_i$$

$$\hat{p}_{i \cdot j} = \hat{p}_i \cdot 1 - d_{ij} / d_i$$

$$\begin{aligned}\hat{q}_{i,j} &= 1 - (1 - \hat{q}_i)^{1-d_{ij}/d_i} \\ &\approx (1 - d_{ij}/d_i) \hat{q}_i\end{aligned}$$

De eerstgenoemde schatter van $q_{i,j}$ is gebaseerd op de onafhankelijke sterftekans, de laatstgenoemde op de afhankelijke sterftekans (zie paragraaf 8.3). De gevonden schatters weerspiegelen de aanname van geheugenloosheid met betrekking tot de opeenvolgende sterfte-aantallen.

Met behulp van de gemeenschappelijke kansgenererende functie kunnen de volgende momenten van de schatters bepaald worden (notatie: $p_{0i} = \Pi(j=0..i-1) p_j$):

$$\begin{aligned}\text{mbt } d_{ij} \quad E(d_{ij} \mid 10) &= 10 * p_{0i} * Q_{ij} \\ \text{var}(d_{ij} \mid 10) &= 10 * p_{0i} * Q_{ij} * (1 - p_{0i} * Q_{ij}) \\ \text{cov}(d_{i1j1}, d_{i2j2} \mid 10) &= -10 p_{0i1} Q_{i1j1} p_{0i2} Q_{i2j2}\end{aligned}$$

$$\begin{aligned}\text{mbt } l_i \quad E(l_i \mid 10) &= 10 * p_{0i} \\ \text{var}(l_i \mid 10) &= 10 * p_{0i} * (1 - p_{0i}) \\ \text{cov}(l_{i1}, l_{i2} \mid 10) &= 10 * (1 - p_{0i1}) * p_{0i2} \text{ (igv } i1 < i2) \\ \text{cov}(d_{i1j}, l_{i2} \mid 10) &= -10 p_{0i1} Q_{i1j} p_{0i2} \\ \text{cov}(l_{i1}, d_{i2j} \mid 10) &= 10 (1 - p_{0i1}) p_{0i2} Q_{i2j}\end{aligned}$$

$$\begin{aligned}\text{mbt } \hat{p}_i, \hat{q}_i \quad \text{var}(\hat{p}_i \mid 10) &= \text{var}(\hat{q}_i \mid 10) \\ &= E(1/l_i) * p_i * q_i = p_i * q_i * p_{0i} / 10 \\ \text{cov}(\hat{p}_{i1}, \hat{p}_{i2} \mid 10) &= \text{cov}(\hat{q}_{i1}, \hat{q}_{i2} \mid 10) = 0 \text{ (igv } i1 \neq i2)\end{aligned}$$

$$\begin{aligned}\text{mbt } \hat{Q}_{ij} \quad \text{var}(\hat{Q}_{ij} \mid 10) &= E(1/l_i) * Q_{ij} * (1 - Q_{ij}) \\ &= Q_{ij} * 1 - Q_{ij} * p_{0i} / 10 \\ \text{cov}(\hat{p}_{i1}, \hat{Q}_{i2j} \mid 10) &= -p_{i1} * Q_{i2j} * p_{0i} / 10 \text{ (igv } i1 = i2) \\ &= 0 \text{ (igv } i1 \neq i2)\end{aligned}$$

Paragraaf 8.7 Uitwerking van de proportional hazards aanname.

De proportional hazards aanname geeft aan, dat de relatieve mortaliteitsdruk voor een specifieke doodsoorzaak ten opzichte van de totale mortaliteitsdruk constant is over een (leef-)tijdsinterval. Omdat uit iedere mortaliteitsdruk de bijbehorende overlevingsfunctie afgeleid kan worden, legt deze veronderstelling dus ook een verband op tussen de overlevingsfuncties voor de afzonderlijke doodsoorzaken en met de algemene overlevingsfunctie.

Uit de proportional hazards aanname volgt (voor willekeurige $j \in J$):

$$S_j(x) = S(x)^{c_j}$$

en voor $p_j = c_j / c_1$ ($j \in J$):

$$S_j(x) = S_1(x)^{p_j}$$

Deze laatste vergelijking geeft aan, dat de functie $S_j(x)$ de overlevingsfunctie beschrijft van het minimum van p_j onderling onafhankelijke stochasten met gelijkvormige overlevingskansen $S_1(x)$. Deze interpretatie is niet alleen toepasbaar voor positieve gehele waarden van p_j , maar ook voor positieve reële waarden. Overigens is de keuze van stochast 1 als referentiefunctie willekeurig. Omdat $c_j > 0$ ($j \in J$), kan iedere stochast deze rol vervullen. In de literatuur worden drie klassen van zogenaamde 'extreme value distributions of the minimum' genoemd, welke aan deze vergelijking voldoen. Deze klassen zijn (voor $a, b > 0$):

klasse I	$S_{a,b}(x) =$	$\exp(-b * (-x)^{-a})$	$x < 0$
		0	$x \geq 0$
II	$S_{a,b}(x) =$	1	$x < 0$
		$\exp(-b * x^a)$	$x > 0$
III	$S_b(x) =$	$\exp(-b * \exp(x))$	

De functies in klasse II zijn de Weibull functies, welke voor $a=1$ overgaan in de exponentiële functies. De proportional hazards aanname laat nog ruimte over voor de toepassing van verschillende vormen van overlevingsfuncties.

Paragraaf 8.3 Toepassing voor Nederland.

Het actuariële model wordt toegepast op de Nederlandse situatie (zie ook paragraaf 1.10). Uitgegaan wordt van de gemiddelde sterfte over de jaren 1986 en 1987 (CBS, 1986, 1987). Van de totale sterfte worden onderscheiden de doodsoorzaken ischaemische hartziekten (ICD 410-414, 420-429), kanker (ICD 140-208) en ongevallen (ICD E801-989). Merk op, dat in dit geval de doodsoorzaak ischaemische hartziekten ruimer gedefinieerd is dan gebruikelijk. De toegepaste berekeningswijzen zijn beschreven in bijlage A en gaan uit van afhankelijke sterftekansen (zie paragraaf 8.3).

In tabel 3 wordt de proportionale verdeling van de sterfte over de drie doodsoorzaken gegeven. Kolom 1 beschrijft de toegepaste verdeling in leeftijdsklassen. Kolom 2 bevat de geslachts- en leeftijdspecifieke algemene sterftecijfers (per 1000), de kolommen 3 t/m 5 de proportionale verdeling daarvan over de genoemde doodsoorzaken.

Tabel 3: De verdeling van de sterfte over de doodsoorzaken.

Leeftijd	Sterfte- fractie (per 1000)	IBZ (in procenten)	Kanker (in procenten)	Ongevallen (in procenten)
Manen				
0	2,15	0,5	2,4	7,1
5	0,22	2,0	18,9	38,8
10	0,22	3,2	18,7	46,0
15	0,51	3,0	10,9	62,1
20	0,75	3,6	10,0	64,3
25	0,73	5,1	12,8	54,7
30	0,83	11,1	17,2	42,0
35	1,18	18,7	24,5	28,7
40	1,90	30,7	29,5	19,9
45	3,30	31,3	34,5	13,0
50	5,81	34,8	39,7	7,4
55	10,22	36,1	41,3	4,4
60	17,32	34,5	40,4	2,8
65	28,83	35,2	40,6	2,3
70	46,70	32,8	36,9	1,9
75	74,96	31,8	32,9	2,1
80	114,16	30,7	27,8	2,2
85	202,89	29,8	19,4	3,1
Totaal:		31,0	31,8	5,1

Proportionale verdeling van verloren levensjaren (in procenten):

27,0	30,4	11,1
------	------	------

Leeftijd	Sterfte- fractie (per 1000)	IHZ (in procenten)	Kanker	Ongevallen
----------	-----------------------------------	-----------------------	--------	------------

Vrouwen

0	1,67	0,7	1,6	4,8
5	0,16	0,7	23,7	21,5
10	0,18	1,7	16,5	36,4
15	0,24	3,2	16,0	46,8
20	0,33	3,0	17,2	47,5
25	0,39	3,2	20,9	40,9
30	0,58	5,5	32,6	29,2
35	0,78	7,8	43,3	21,1
40	1,24	9,0	58,3	15,8
45	2,08	10,0	56,2	12,0
50	3,24	14,0	58,2	6,5
55	5,17	16,6	55,5	5,5
60	7,82	20,8	49,6	3,4
65	12,26	27,2	43,5	2,6
70	20,69	31,9	34,3	2,4
75	37,31	34,2	25,2	2,3
80	68,19	34,2	20,2	2,6
85	154,30	32,3	14,2	3,6

Totaal:		29,2	26,7	4,2
---------	--	------	------	-----

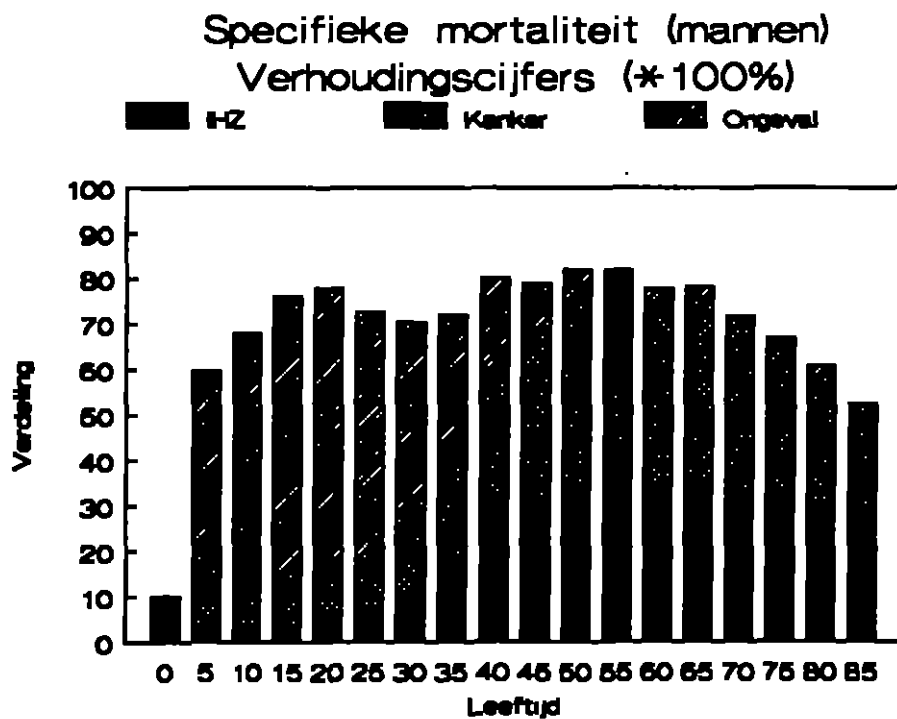
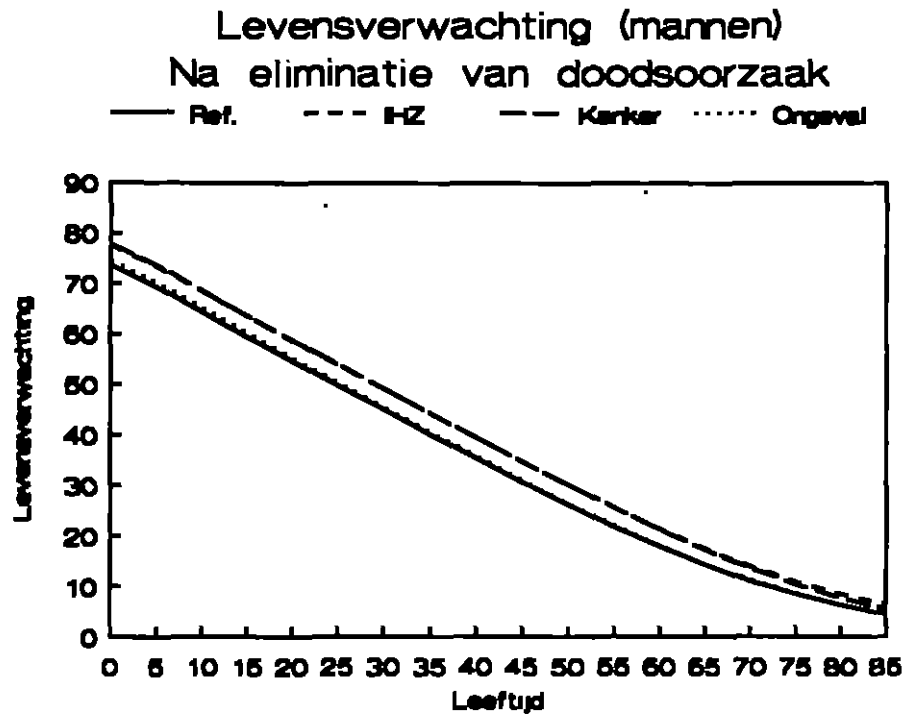
.. Proportionele verdeling van verloren levensjaren (in procenten):

	22,6	32,2	7,3
--	------	------	-----

Bron: CBS, 1986, 1987

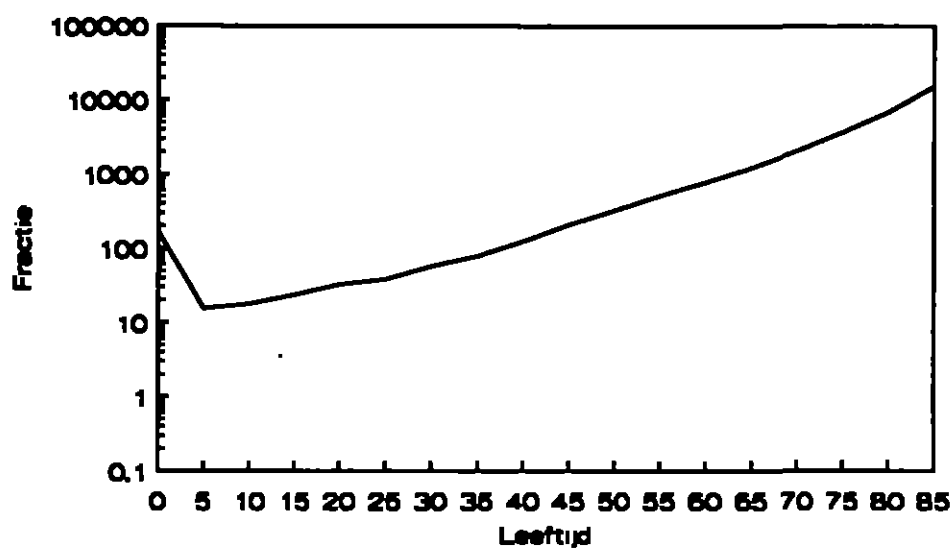
Tevens is in tabel 3 de proportionele verdeling over de drie genoemde doodsoorzaken gegeven van het totaal aantal verloren levensjaren ten gevolge van sterfte. Om meer inzicht in de mortaliteitsverschillen tussen de drie specifieke doodsoorzaken te geven zijn voor achtereenvolgens mannen en vrouwen de sterftefracties en de verdeling daarvan over de drie doodsoorzaken ook grafisch weergegeven (zie de figuren 12 en 13).

Vervolgens zijn drie sterftetabellen-(tabel 4) geconstrueerd na eliminatie van de doodsoorzaken IHZ resp. kanker resp. ongevallen in het actuariële model. Deze sterftetabellen zijn minder uitgebreid dan de basis-sterftetabel (zie paragraaf 1.9) en bevatten slechts de overlevingsaantallen en de resterende levensverwachting. Met name de levensverwachting toont duidelijk de verschillen tussen de invloed van de verschillende doodsoorzaken. Vergelijking met tabel 1 (paragraaf 1.9) laat zien, dat eliminatie van ongevallen als doodsoorzaak slechts een zeer geringe toename van de (resterende) levensverwachting tot gevolg heeft, en dat de grootste toename van de resterende levensverwachting na eliminatie van IHZ of kanker optreedt op middelbare leeftijd.



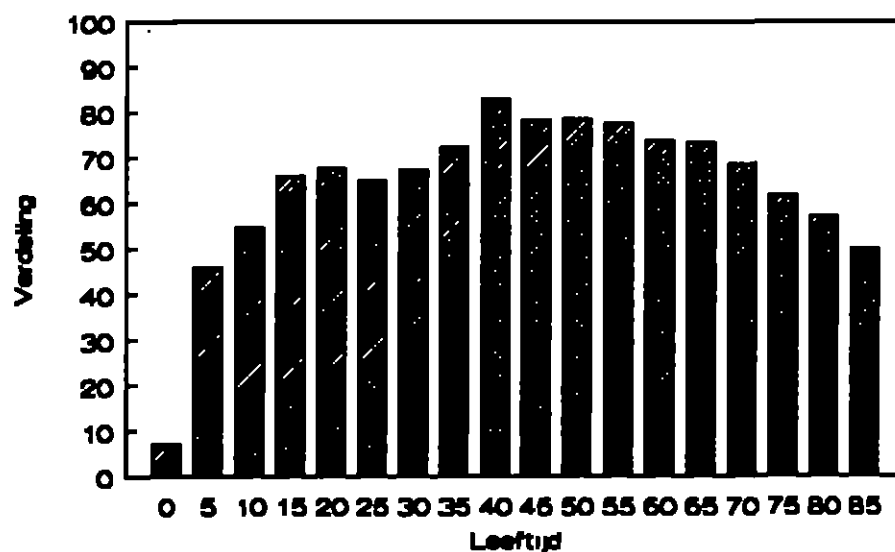
Figuur 12: de doodsoorzaak- en leeftijd-specifieke sterftefracties voor mannen, absoluut en relatief.

Sterftefracties (vrouwen) Per 100.000



Specifieke mortaliteit (vrouwen) Verhoudingscijfers (*100%)

■ I-Z ■ Kanker ■ Ongeval

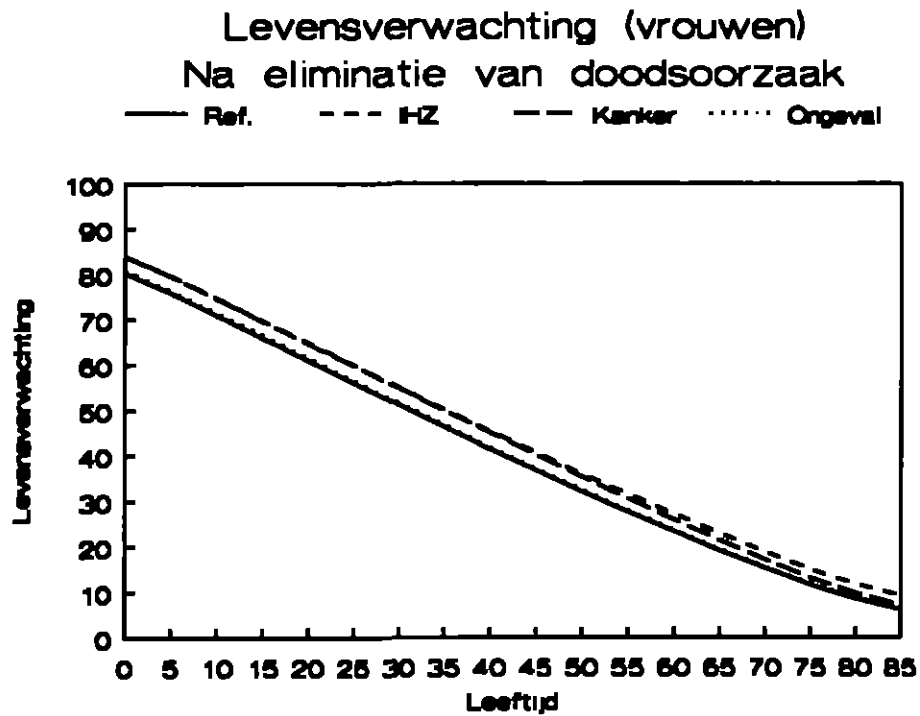
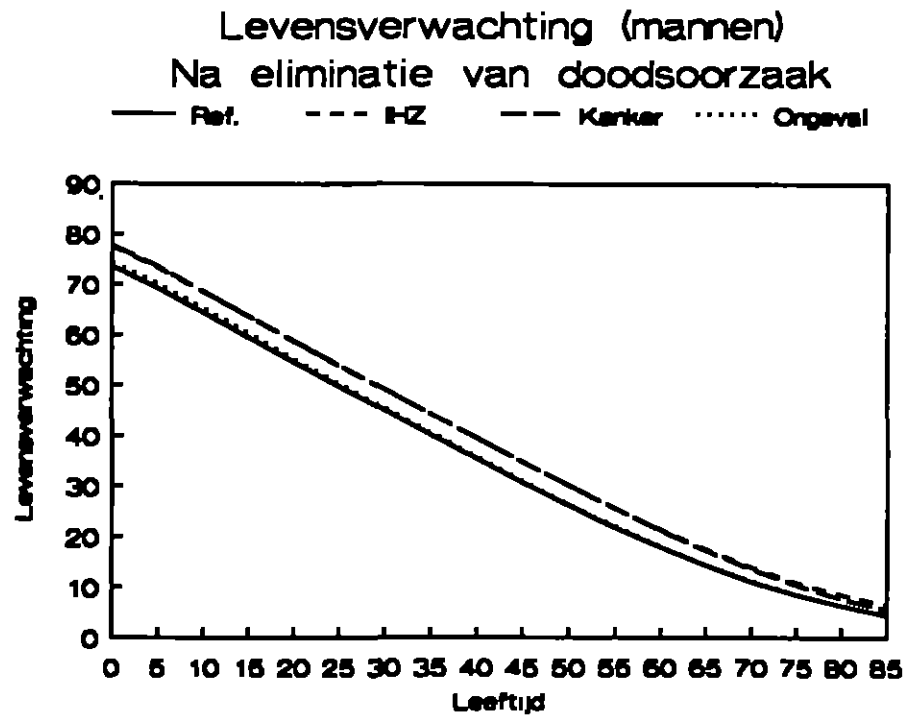


Figuur 13: de doodsoorzaak- en leeftijd-specifieke sterftefracties voor vrouwen, absoluut en relatief.

Tabel 4: De sterftetabellen na eliminatie van een doodsoorzaak.

Leeftijd Geen ICH			Geen kanker		Geen ongevallen	
	Overlev. aantallen	Levens- verwacht.	Overlev. aantallen	Levens- verwacht.	Overlev. aantallen	Levens- verwacht.
Mannen						
0	100.000	77,6	100.000	77,8	100.000	74,5
5	98.935	73,4	98.955	73,6	99.006	70,2
10	98.831	68,5	98.869	68,6	98.940	65,3
15	98.728	63,6	98.783	63,7	98.883	60,3
20	98.484	58,7	98.558	58,8	98.787	55,3
25	98.131	53,9	98.229	54,0	98.565	50,4
30	97.793	49,1	97.918	49,2	98.494	45,5
35	97.433	44,3	97.582	44,3	98.257	40,6
40	96.967	39,5	97.148	39,5	97.844	35,8
45	96.331	34,7	96.501	34,8	97.104	31,0
50	95.245	30,1	95.464	30,1	95.721	26,4
55	93.456	25,6	93.804	25,6	93.175	22,1
60	90.442	21,4	91.025	21,3	88.713	18,1
65	85.431	17,5	86.427	17,3	81.492	14,4
70	77.749	14,0	79.279	13,6	70.640	11,3
75	66.285	10,9	68.268	10,4	55.878	8,6
80	50.984	8,5	52.740	7,8	38.167	6,4
85	33.762	6,5	34.295	5,6	21.111	4,6
Vrouwen						
0	100.000	84,0	100.000	83,9	100.000	80,8
5	99.159	79,7	99.167	79,6	99.194	76,5
10	99.083	74,8	99.108	74,6	99.133	71,5
15	98.995	69,8	99.034	69,7	99.077	66,6
20	98.883	64,9	98.936	64,7	99.015	61,6
25	98.727	60,0	98.803	59,8	98.930	56,6
30	98.543	55,1	98.653	54,9	98.818	51,7
35	98.275	50,2	98.462	50,0	98.617	46,8
40	97.923	45,4	98.245	45,1	98.314	41,9
45	97.371	40,7	97.991	40,2	97.802	37,1
50	96.465	36,1	97.546	35,4	96.912	32,5
55	95.130	31,5	96.888	30,6	95.455	27,9
60	93.098	27,1	95.778	25,9	93.145	23,6
65	90.252	22,9	93.908	21,4	89.680	19,4
70	86.297	18,8	90.703	17,1	84.452	15,4
75	80.385	15,0	84.704	13,1	76.262	11,8
80	70.994	11,7	73.530	9,7	63.341	8,7
85	56.435	9,1	55.577	7,1	44.922	6,2

..Bron: CBS, 1986, 1987



Figuur 14: De levensverwachting na eliminatie van een doodsoorzaak.

De leeftijds specifieke levensverwachting na eliminatie van een doodsoorzaak, zoals beschreven in tabel 4, is grafisch weergegeven in figuur 14.

Paragraaf 8.9 Afsluiting.

Het actuariële model biedt een eerste mogelijkheid van uitwerking van het concept concurrerende doodsoorzaken. Via invulling van de specifieke sterfteaantallen aan de hand van bestaande sterfteregistraties kunnen de waarden van de meest aannemelijke schatters van de verschillende overlevings- en sterftekansen bepaald worden. Aan de hand van deze uitkomsten kan vervolgens een eerste kwantitatieve beschrijving gegeven worden van de relatie tussen bijvoorbeeld (een verlaging van) de sterfte aan hart- en vaatziekten en (een verhoging van) de sterfte aan kanker.

Deze aanpak heeft echter weinig realiteitswaarde, indien een verlaging van de sterfte aan hart- en vaatziekten gerealiseerd wordt met (bijvoorbeeld) behulp van preventiemaatregelen. Deze zijn bedoeld om het niveau van één of meer risicovariabelen te beïnvloeden en daarmee één of meer doodsoorzaak-specifieke mortaliteitsdrukken. De veronderstelling, dat de doodsoorzaken vervangend zijn, is dan weinig realistisch en dient zo mogelijk voor een andere aanpak gekozen te worden.

HOOFDSTUK 9 Geregistreerde sterftcijfers en overlevingsfuncties.

Paragraaf 9.1 Inleiding.

In hoofdstuk 1 is een theoretische uiteenzetting van overlevingsfuncties gegeven. Deze overlevingsfuncties beschrijven voor de achtereenvolgende leeftijdsjaren de overlevingsfracties van een cohort. Uit deze overlevingsfuncties c.q. -fracties kan een aantal andere grootheden afgeleid worden. Alle genoemde grootheden werden (voorlopig) continu afhankelijk van de leeftijd verondersteld. Geregistreerde sterftcijfers worden daarentegen discreet per leeftijdsklasse gegeven. De centrale sterftegraad blijkt een belangrijke rol te spelen bij het bepalen van de waarden van de overlevingsfracties en daarvan afgeleide grootheden op basis van deze geregistreerde cijfers.

Daarnaast is bij geregistreerde sterftcijfers de waarnemingstijd van belang, en wel langs twee wegen. De genoemde overlevingsfuncties en daarvan afgeleide grootheden zijn niet slechts afhankelijk van de leeftijd, maar ook van de waarnemingstijd. Zo zijn de leeftijds-specifieke sterftcijfers de laatste tientallen jaren gedaald. Daarnaast volgt de momentane frequentieverdeling van een bevolking niet de (cohort-)overlevingsfunctie, maar is het resultaat van de geboortecijfers, de migratiecijfers en de leeftijd-afhankelijke overlevingsfracties uit het verleden. Een duidelijk voorbeeld van het verschil tussen overlevingsfracties en de frequentieverdeling vormt de doorwerking van de geboortegolf van na de tweede wereldoorlog.

In dit hoofdstuk zullen onder verschillende aannames voorwaardelijke overlevingsfracties afgeleid worden uit geregistreerde sterftcijfers (zie bijvoorbeeld Keyfitz, 1968). In eerste instantie wordt verondersteld, dat de waarnemingstijd slechts langs de laatstgenoemde weg een rol speelt: de frequentieverdeling en de overlevingsfracties komen niet overeen. De volgende grootheden worden nu toegepast:

$l(x)$	de (onvoorwaardelijke) overlevingsfractie op leeftijd x
$L(x)$	de geaccumuleerde overlevingsfracties tussen de leeftijden x en $x+5$,
$n(x)$	de momentane frequentieverdeling van de bevolking verdeeld naar leeftijd
$N(x)$	de momentane frequentieverdeling van de bevolking verdeeld naar leeftijdsinterval,

$D(x)$	de geregistreerde eenjaars sterfte onder het bovengenoemde aantal $N(x)$
$M(x)$	de centrale sterftegraad tussen leeftijd x en $x+5$
$h(x)$	de mortaliteitsdruk
$p(x,x+5)$	de 5-jaars overlevingskans op leeftijd x

Tussen deze grootheden bestaan de volgende verbanden:

$$\begin{aligned}
 L(x) &= l(x,x+5) = \int_x^{x+5} l(s) \, ds \\
 N(x) &= n(x,x+5) = \int_x^{x+5} n(s) \, ds \\
 M(x) &= D(x) / N(x) = \int_x^{x+5} n(s) h(s) \, ds / \int_x^{x+5} n(s) \, ds \\
 p(x,x+5) &= l(x+5) / l(x) = \exp\left(- \int_x^{x+5} h(s) \, ds\right)
 \end{aligned}$$

- De centrale sterftegraad kan beschouwd worden als een gewogen gemiddelde van de mortaliteitsdruk. De voorwaardelijke overlevingsfractie is een functie van de cumulatieve mortaliteitsdruk.

Paragraaf 9.2 De voorwaardelijke overlevingsfractie als functie van de sterftegraad voor een stabiele bevolking.

Onder verschillende aannames kan de voorwaardelijke overlevingsfractie p geschreven worden als functie van de sterftegraad m . In deze paragraaf worden drie voorbeelden van aannames gegeven, die in zekere mate een stabiele opbouw van de bevolking veronderstellen. In de volgende paragraaf wordt een groeiende bevolking verondersteld. De drie voorbeelden van aannames onder een stabiel veronderstelde bevolking zijn:

Constante mortaliteitsdruk

Dan geldt:

$$\begin{aligned}
 M(x) &= 1 - \exp(-cst) \\
 p(x,x+5) &= \exp(-5 * cst) = (1 - M(x))^5 \quad (\text{oplossing 1}) \\
 &= \exp(-5 * M(x)) \quad (\text{oplossing 2})
 \end{aligned}$$

Frequentieverdeling en overlevings aantallen zijn identiek en lineair

Dan geldt:

$$\begin{aligned}
D(x) &= \int_x^{x+5} n(s) h(s) ds = \int_x^{x+5} l(s) h(s) ds = l(x) - l(x+5) \\
M(x) &= 2/5 * l(x) - l(x+5) / l(x) + l(x+5) \\
p(x, x+5) &= \frac{1 - \frac{5}{2} \frac{M(x)}{l(x)}}{1 + \frac{5}{2} \frac{M(x)}{l(x)}} \quad (\text{oplossing 3})
\end{aligned}$$

Frequentieverdeling en mortaliteitsdruk zijn lineair.

Dan geldt:

$$\begin{aligned}
N(x) &= \int_x^{x+5} n(s) ds = 5 * n(x, x+5/2) \\
D(x) &= \int_x^{x+5} n(s) h(s) ds \\
&= 5 n(x+5/2) h(x+5/2) + 125/12 n'(x+5/2) h'(x+5/2) \\
n'(x+5/2) &= (N(x+5) - N(x-5)) / 2 * 5 / 5 \\
h'(x+5/2) &= (M(x+5) - M(x-5)) / 2 / 5 \\
p(x, x+5) &= D(x) / N(x) = \exp(-5 (M(x) + c)) \\
\text{met } c &= \frac{(N(x+5) - N(x-5))}{48 N(x)} * \frac{(M(x+5) - M(x-5))}{2} \quad (\text{oplossing 4})
\end{aligned}$$

De laatste oplossing komt op de correctiefactor na overeen met de tweede oplossing.

Van de gevonden oplossingen kunnen de volgende orde-benaderingen gegeven worden (voor $m = M(x)$):

oplossing	$p(x, x+5)$
1	$(1 - m)^5 = 1 - 5m + 10m^2 - 10m^3 + 0(m^4)$
2	$\exp^{-5m} = 1 - 5m + 12.5m^2 - 13.9m^3 + 0(m^4)$
3	$(1 - 5m/2)/(1 + 5m/2) = 1 - 5m + 12.5m^2 - 31m^3 + 0(m^4)$
4	$\exp^{-5m+c} = 1 - 5m+c + 12.5m+c^2 - 13.9m+c^4 + 0(m+c^4)$

Hieruit blijkt, dat de eerste drie oplossingen in de eerste orde gelijk zijn en daarvan de laatste twee bovendien ook in de tweede orde. De vierde oplossing wijkt vanaf de eerste orde reeds af van de overige oplossingen.

Paragraaf 9.3 De voorwaardelijke overlevingskans als functie van de sterftegraad voor een groeiende bevolking.

In de vorige paragraaf zijn drie voorbeelden van aannames voor een min of meer constante bevolkingsopbouw uitgewerkt. In deze paragraaf wordt de voorwaardelijke overlevingskans berekend onder de aanname van een constant groeiende bevolking. In eerste instantie wordt de exponentiële groeifactor r bekend verondersteld. De verschillende vergelijkingen worden nu:

$$\begin{aligned}
 n(x) &= \text{cst} * e^{-r*x} * l(x) \\
 D(x) &= \int_x^{x+5} n(s) h(s) ds = - \int_x^{x+5} e^{-rs} l'(s) ds \\
 &= -r * \int_x^{x+5} n(s) ds - \int_x^{x+5} n'(s) ds \\
 &= -r * N(x) + (n(x) - n(x+5)) \\
 N(x) &= \int_x^{x+5} n(s) ds \\
 &= \frac{65}{24} * (n(x) + n(x+5)) - \frac{5}{24} * (n(x-5) + n(x+10)) \\
 M(x) + r &= \frac{65/24 * (n(x) + n(x+5)) - 5/24 * (n(x-5) + n(x+10))}{n(x)} \\
 p(x, x+5) &= l(x+5) / l(x) \\
 &= \exp^{5r} \frac{1 - (M(x)+r) \left| \frac{65}{24} - \frac{5}{24} e^{-10r} l(x+10) + e^{5r} l(x-5) \right|}{1 + (M(x)+r) * 65/24}
 \end{aligned}$$

De laatste vergelijking representeert een stelsel vergelijkingen van de leeftijds specifieke voorwaardelijke overlevingskansen voor gegeven groeifactor r en sterftegraad m . Uit dit stelsel vergelijkingen en de voorafgaande vergelijking, waarin de sterftegraad als functie van de groeifactor en de overlevingskansen beschreven wordt, kunnen beide onbekenden iteratief berekend worden.

De (onbekende) exponentiële groeifactor r beschrijft de relatie tussen overlevingsaantallen l en frequentieverdeling n . Deze groeifactor dient daarom ook tezamen met de overlevingskansen l en sterftegraad m middels het genoemde iteratieve rekenproces bepaald te worden. Voor de bepaling van de groeifactor en daarmee van de frequentieverdeling tijdens een stap uit dit rekenproces moeten (minimaal) twee leeftijdsintervallen meegenomen worden. Het ligt uit symmetrie overwegingen voor de hand om hiervoor de beide buurintervallen van het interval $[x, x+5)$ te kiezen. Twee oplossingen zijn nu mogelijk, waarvan de tweede meer verkiend is.

Oplossingsmethode 1.

Hierbij worden de aantallen van de frequentieverdeling gecentreerd in het midden van het leeftijdsinterval verondersteld, ofwel:

$$N(x-5) = \text{cst} * \exp^{-r * x-2.5} * L(x-5)$$

$$N(x+5) = \text{cst} * \exp^{-r * x+7.5} * L(x+5)$$

waaruit volgt:

$$r = (\ln N(x-5) - \ln L(x-5) - \ln N(x+5) + \ln L(x+5)) / 10$$

Oplossingsmethode 2.

De tweede methode vormt een verfijning van de eerste. Als uitgangspunt wordt de volgende vergelijking gekozen:

$$\frac{N(x+5)}{N(x-5)} = \frac{\int \frac{n(x+5+s)}{n(x-5+s)} ds}{\int \frac{n(x-5+s)}{n(x-5+s)} ds} = \frac{\int \frac{\exp(-r(x+5+s)) * 1(x+5+s)}{\exp(-r(x-5+s)) * 1(x-5+s)} ds}{\int \frac{\exp(-r(x-5+s)) * 1(x-5+s)}{\exp(-r(x-5+s)) * 1(x-5+s)} ds}$$

De beschreven integralen worden over het interval [0,5) genomen. Worden in deze vergelijking slechts de eerste orde invloeden van de groeifactor r meegenomen, dan kan factor benaderd worden met behulp van de formule:

$$r = \frac{(\ln N(x-5) - \ln L(x-5) - \ln N(x+5) + \ln L(x+5)) / 10}{1 + \left(\frac{L(x-10) - L(x)}{L(x-5)} - \frac{L(x) - L(x+10)}{L(x+5)} \right) / 48}$$

De teller van deze breuk komt overeen met de oplossing volgens de eerste methode, de noemer bevat een correctieterm.

HOOFDSTUK 10 Conclusies en aanbevelingen.

Paragraaf 10.1 Inleiding.

Het onderhavige rapport vormt het verslag van een literatuurstudie op het terrein van overlevingsduuranalyse en concurrerende doodsoorzaken. De nadruk heeft bij de literatuurstudie gelegen op de wiskundig-statistische aspecten van het onderwerp. Zo is bij de beschrijving van het risico op sterfte (hoofdstuk 5) de nadruk gelegd op de verschillende mogelijkheden van wiskundige modellering van het risico. Epidemiologische aspecten of uitwerkingen van de wiskundige methoden voor specifieke voorbeelden zijn niet beschreven. Daarvoor wordt verwezen naar (de aangehaalde) artikelen in statistische en epidemiologische tijdschriften.

De nadruk op de wiskundig-statistische achtergronden van overlevingsduuranalyse en concurrerende doodsoorzaken maakt dat het rapport geen handboek vormt voor het toepassen van overlevingsduuranalyse of een analyse van concurrerende doodsoorzaken op praktische probleemstellingen. Van de lezer wordt daarom een zekere expertise op het onderzoeksterrein verwacht om (onderdelen van) de inhoud van het rapport toe te kunnen passen.

Paragraaf 10.2 Enkele conclusies op basis van de studie.

Op basis van de uitgevoerde literatuurstudie kunnen conclusies getrokken worden, die betrekking hebben op de invloed van concurrerende doodsoorzaken en/of op de Nederlandse situatie. Deze conclusies worden in deze paragraaf beschreven. Daarnaast kunnen conclusies getrokken en aanbevelingen geformuleerd worden, die gericht zijn op de mogelijke toepassingen van overlevingsduuranalyse en concurrerende doodsoorzaken, zoals genoemd in het Voorwoord. Deze conclusies en aanbevelingen worden beschreven in de volgende paragraaf.

De eerste conclusie, die op basis van de literatuurstudie getrokken kan worden, luidt (zie de hoofdstukken 6 tot en met 8):

- Doodsoorzaken met gemeenschappelijke belangrijke risicofactoren dienen als concurrerend in plaats van vervangend beschouwd te worden.

Deze conclusie is met name van belang, indien preventiemaatregelen toegepast worden om de sterfte aan specifieke doodsoorzaken terug te dringen. Deze preventiemaatregelen zijn bedoeld om het risico op sterfte aan specifieke doodsoorzaken te verlagen. Indien de te

veranderen risicofactoren van invloed zijn op verschillende doodsoorzaken is een epidemiologische aanpak gewenst.

Bijvoorbeeld: vermindering van de sigaretten-consumptie leidt enerzijds tot een vermindering van de sterfte aan ischaemische hartziekten en longkanker (in strikte zin concurrerende doodsoorzaken, omdat de sigarettenconsumptie een gemeenschappelijke determinant vormt). Vanwege deze vermindering van de sterfte overleven meer mensen, wat anderzijds leidt tot een vergroting van de sterfte aan longkanker respectievelijk ischaemische hartziekten (vervangende doodsoorzaken). Het netto-effect van beide ontwikkelingen is niet eenvoudig te beschrijven, temeer omdat tijdeffecten een belangrijke rol spelen (zie Gunning-Schepers, 1988).

De tweede conclusie luidt (zie hoofdstuk 4):

- De aanwezigheid van belangrijke risicogroepen leidt tot een heterogeniteit van de bevolking en beïnvloedt de sterfte.

Deze conclusie is met name van belang, indien preventiemaatregelen gericht op specifieke risicogroepen (doelgroepen) toegepast worden. Bijvoorbeeld: wiskundige modellen van de sterfte aan AIDS, die gebaseerd zijn op een homogene of heterogene bevolking, vertonen in de loop van de tijd duidelijke verschillen.

De derde conclusie luidt (zie hoofdstuk 8):

- In Nederland is de afgelopen tientallen jaren voor lage leeftijden de resterende levensverwachting duidelijk toegenomen; voor hoge leeftijden vrijwel niet.

Deze conclusie kan ook anders geformuleerd worden: de overlevings-curve van de Nederlandse bevolking wordt steeds meer rechthoekig van vorm. Vanwege de geringe sterfte op lage leeftijd leidt (bijvoorbeeld) een verdere vermindering van de ongevalssterfte vrijwel niet tot een verlenging van de gemiddelde levensverwachting.

De vierde conclusie luidt (zie hoofdstuk 8):

- Het belang van een doodsoorzaak en daarmee van een maatregel om de sterfte aan deze doodsoorzaak te verminderen wordt sterk bepaald door het toegepaste criterium.

Deze conclusie is getrokken op basis van een vergelijking van de invloed van de doodsoorzaken ischaemische hartziekten, kanker en ongevallen aan de hand van de criteria 'sterfte-aantallen' en 'verloren levensjaren'. Bijvoorbeeld: onder het tweede criterium vormt de ongevalssterfte een veel belangrijker doodsoorzaak dan onder het eerste. De genoemde criteria hebben betrekking op mortaliteit.

Daarnaast kunnen criteria onderscheiden worden gericht op morbiditeit en kwaliteit van leven (zie Hoogenveen, Verkleij e.a.,

1990). In dit kader kunnen genoemd worden: de prevalentie (beschrijft het voorkomen van een ziekte); de ADL-schaal (Activiteiten van het Dagelijks Leven, beschrijft de (in)validiteit en mobiliteit van personen); de VOG-score (Vragenlijst Over de Ervaren Gezondheids- toestand, beschrijft de subjectieve gezondheidsbeleving), en de Quality Adjusted Life Years (Qaly's, beschrijft de utiliteitswaarde van gezondheidstoestand en kwaliteit van leven). Bijvoorbeeld: chronische, niet-letale ziekten zijn onder criteria gericht op ziekte en gevolgen van ziekte van veel groter belang dan onder criteria gericht op mortaliteit (zie ook Colvez, Blanchet, 1983).

De vijfde conclusie luidt:

- Met betrekking tot (trends in) sterfte zijn relatief veel data verzameld (maar daarom nog niet eenvoudig beschikbaar), met betrekking tot trends in risico en het verband tussen sterfte en risico duidelijk minder.

Zoals uit de eerste en tweede conclusie naar voren komt, speelt bij een onderzoek naar concurrerende doodsoorzaken de invloed van risico-factoren op (doodsoorzaak-specifieke) sterfte een belangrijke rol. Cijfers over trends in risicofactoren en over relatieve risico's zijn daarom noodzakelijk. Gegevensverzamelingen, die trends in risicofactoren beschrijven, zijn over het algemeen slechts fragmentarisch, kort-lopend en regio-bepakt. Indien het risico multiplicatief verondersteld wordt en onder aanname van onderlinge onafhankelijkheid van risicofactoren kunnen ook (gestandaardiseerde) relatieve risico's, verkregen uit onderzoeken van buiten Nederland, gebruikt worden (zie Gunning-Schepers, 1988).

Paragraaf 10.3 Conclusies en aanbevelingen met betrekking tot vervolgonderzoeken.

In de literatuurstudie, die aan het onderhavige rapport ten grondslag ligt, is met name gezocht naar literatuur over de wiskundig-statistische achtergronden van overlevingsduuranalyse en concurrerende doodsoorzaken. Aanbevelingen voor verder onderzoek zijn niet rechtstreeks uit deze studie te destilleren, maar dienen tegen de achtergrond van bestaande vraagstellingen opgesteld te worden.

Tot deze 'achtergrond' (zie het Voorwoord) kunnen de volgende onderzoeken resp. documenten gerekend worden. Ten eerste: het scenario-onderzoek, dat uitgevoerd wordt onder auspiciën van de Stuurgroep Toekomstscenario's Gezondheidszorg. Aan verschillende onderwerpen op het terrein van de volksgezondheid en gezondheidszorg zijn afzonderlijke scenario-projecten gewijd, waarin door scenario-commissies en bijbehorende onderzoeksgroepen toekomstverkenningen opgesteld zijn c.q. worden. Ten tweede: het Ontwerp-Kerndocument Gezondheidsbeleid, waarin een aantal gezondheidsdoelen voor

toekomstig beleid geformuleerd wordt. Ten derde: het door het RIVM uit te brengen 'tweede Zorgen voor Morgen', waarin ook de invloed van het milieu op de volksgezondheid beschreven wordt.

De eerste conclusie.

Deze conclusie heeft betrekking op belangrijke problemen, die genoemd kunnen worden bij het wiskundig-theoretisch onderzoek van concurrerende doodsoorzaken en waarvoor een oplossing gevonden moet worden. In het rapport is onderscheid gemaakt tussen vervangende en concurrerende doodsoorzaken (zie paragraaf 7.9). Doodsoorzaken worden vervangend genoemd, indien een verandering van de mortaliteitsdruk voor een willekeurige doodsoorzaak niet gepaard gaat met een verandering van de mortaliteitsdruk voor de overige doodsoorzaken. Doodsoorzaken heten concurrerend, indien ze niet vervangend zijn. Bij dit onderscheid tussen vervangende en concurrerende doodsoorzaken speelt het begrip gemeenschappelijke risicofactor (determinant) een essentiële rol.

Bij vervangende doodsoorzaken wordt onderscheid gemaakt tussen onafhankelijke en afhankelijke sterftekanen (zie paragraaf 8.3). Bij afhankelijke sterftekanen wordt ervan uitgegaan, dat de doodsoorzaak-specifieke sterftekanen beschreven wordt door de waargenomen relatieve sterfte. Bij onafhankelijke sterftekanen daarentegen wordt rekening gehouden met de concurrentie tussen doodsoorzaken en wordt gecorrigeerd voor de overige sterfte. Uit de eigen onderzoeksresultaten komt naar voren, dat voor de Nederlandse situatie de verschillen klein zijn. In de literatuur wordt hierover ook anders geoordeeld (Van Ginneken e.a., 1989).

Bij concurrerende doodsoorzaken worden doodsoorzaak-specifieke sterftekanen geschreven als functie van één of meer deels gemeenschappelijke risicofactoren (determinanten). Deze risicofactoren kunnen onderling onafhankelijk en afhankelijk verondersteld worden. Maar de mening van de auteurs dient bij verder wiskundig-theoretisch onderzoek van concurrerende doodsoorzaken aan met name het laatstgenoemde geval van concurrerende doodsoorzaken en één of meer deels gemeenschappelijke, onderling afhankelijke risicofactoren aandacht geschonken te worden.

De tweede conclusie.

Deze conclusie heeft betrekking op de beschrijving en waardering van doodsoorzaken in termen van verminderde gezondheid. De invloed van concurrerende doodsoorzaken is in dit rapport slechts beschreven in termen van mortaliteit (tijdstip van overlijden en doodsoorzaak). De invloed van doodsoorzaken wordt daarnaast steeds meer beschreven in termen van morbiditeit en kwaliteit van leven. De belangrijkste achterliggende idee is, dat gezondheidswinst in Nederland meer in termen van kwaliteit van leven dan van mortaliteit te behalen valt. Het is daarom zinvol om ook de genoemde concepten morbiditeit en kwaliteit van leven te beschouwen bij vervolgonderzoeken op het terrein van concurrerende doodsoorzaken.

Daarnaast is het belangrijk om expliciete modelbeschrijvingen te ontwikkelen van ziekteprocessen, die eventueel voorafgaan aan sterfte. Vanwege de toenemende vergrijzing in Nederland neemt het belang van chronische ziekten, die niet direct levensbedreigend zijn, toe. Bij de (beoogde) terugdringing van belangrijke doodsoorzaken als hart- en vaatziekten en kanker dient hiermee rekening gehouden te worden. Een aanzet voor de modellering hiervan vormt het Prevent-model, waarin voor de Nederlandse situatie voor een aantal doodsoorzaken in de tijd de incidentie en sterfte beschreven zijn als functie van het niveau van verschillende gemeenschappelijke risicofactoren (Gunning-Schepers, 1988).

De derde conclusie.

De derde conclusie heeft betrekking op de invulling van de twee bovengenoemde conclusies met betrekking tot het wiskundig-methodologisch onderzoek ten behoeve van de Stuurgroep Toekomst-scenario's Gezondheidszorg en met betrekking tot het Ontwerp-Kerndocument Gezondheidsbeleid. In beide gevallen wordt aandacht besteed aan de doodsoorzaken hart- en vaatziekten (ischaemische hartziekten), kanker, ongevallen en chronische ziekten. In het kader van het STG-onderzoek is een belangrijke invulling het vergelijken en combineren van de in de betreffende scenario-projecten ontwikkelde toekomstscenario's. Met betrekking tot het Ontwerp-Kerndocument bestaat een belangrijke invulling uit het onderzoeken van de onderlinge afhankelijkheid van de verschillende gestelde gezondheidsdoelen. Deze activiteiten dienen in samenhang plaats te vinden.

LITERATUURLIJST

- Aalen O.O. Heterogeneity in survival analysis. *Statistics in Medicine*, vol 7 (1988), pp 1121-37
- Aalen O.O. A linear regression model for the analysis of life times. *Statistics in Medicine*, vol 8 (1989), pp 907-25
- Abbott R.D. Logistic regression in survival analysis. *American Journal of Epidemiology*, vol 121 (1985), pp 465-71
- CBS, Maandbericht van de Gezondheidsstatistiek
- CBS, Overledenen naar doodsoorzaak, leeftijd en geslacht, serie A1
- CBS, Statistisch zakboek
- Chiang C.L. Introduction to stochastic processes in biostatistics. John Wiley & Sons, Inc., New York (etc.), 1968
- Colvez A., Blanchet M. Potential gains in life expectancy free of disability: a tool for health planning. *International Journal of Epidemiology*, vol 12 (1983), pp 224-9
- Cox D.R. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Ser. B*, vol 34 (1972), pp 187-202
- Cox D.R. Partial likelihood. *Biometrika*, vol 62 (1975), pp 269-76
- Cox D.R. Oakes D. Analysis of survival data. Monographs on Statistics and Applied Probability. Chapman and Hall, London, New York, 1984
- David H.A. On Chiang's proportionality assumption in the theory of competing risks, 282 Note. *Biometrics*, vol 26 (1970), pp 767-76
- Efron B. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, vol 72 (1977), pp 557-65
- Follmann D.A., Goldberg M.S. Distinguishing heterogeneity from decreasing hazard rates. *Technometrics*, vol 30 (1989), pp 389-96
- Gail M. A review and critique of some models used in competing risk analysis. *Biometrics*, vol 31 (1975), pp 209-22
- Gill R.D. Understanding Cox's regression model:-a martingale approach. *Journal of the American Statistical Association*, vol 79 (1984), pp 441-7

- Gimeken J.K.S.van, Barmenberg A.F.I. Dissevelt A.G.
Gezondheidsverlies ten gevolge van een aantal belangrijke
ziekte categorieën in 1981 - 1985: methodologische aspecten en
resultaten. Definitief concept-rapport, Fase 1, Deel 1,
Preventieprofielen Project. NIPG Leiden, CBS Voorburg, 1989
- Gurning-Schepers L. The health benefits of prevention, a simulation
approach. Academisch proefschrift. Instituut Maatschappelijke
Gezondheidszorg, Erasmus Universiteit Rotterdam, 1988
- Hempenius A.L., Mulder P.G.H. Multiple failure rates and
observations of time dependent covariables, Part I: Theory.
Research memorandum. Tilburg University, Department of
Economics, 1981
- Hensler G.L., Mehrotra K.G., Michalek J.E. A note on some exact
efficiency calculations for survival distributions. *Biometrika*,
vol 64 (1977), pp 635-7
- Hoogenveen R.T., Verkleij H., Jager J.C., Kromhout D. Verslag van
'Verkennde studie betreffende concurrerende doodsoorzaken in
verband met kwantitatieve gezondheidsdoelen' ten behoeve van de
afdeling Epidemiologie en Informatica van het Ministerie van
VVC. RIVM, Bilthoven, 1990 (in voorbereiding)
- Hougaard P. Life table methods for heterogeneous populations:
Distributions describing the heterogeneity. *Biometrika*, vol 71
(1984), pp 75-83
- Hougaard P. Survival models for heterogeneous populations derived
from stable distributions. *Biometrika*, vol 73 (1986), pp 387-96
- Kalbfleisch J.D., Lawless J.F. Estimation of reliability in field-
performance studies. *Technometrics*, vol 30 (1988), pp 365-88
- Keyfitz N. A life table that agrees with the data. *Journal of the
American Statistical Association*, vol 61 (1966), pp 305-11
- Keyfitz N. A life table that agrees with the data: II. *Journal of
the American Statistical Association*, vol 63 (1968), pp 1253-68
- Keyfitz N. *Applied mathematical demography*. John Wiley & Sons,
London (etc.), 1977
- Keyfitz N., Beckman J.A. *Demography through problems*. Springer
Verlag, New York (etc.), 1984
- Larson M.G., Dinse G.E. A mixture model for the regression analysis
of competing risks data. *Applied Statistics*, vol 34 (1985), pp
201-11
- Link W.A. A model for informative censoring. *Journal of the American
Statistical Association*, vol 84 (1989), pp 749-52

- Manton K.G., Patrick C.H., Stallard E. Population impact of mortality reduction: The effects of elimination of major causes of death on the 'saved' population. *International Journal of Epidemiology*, vol 9 (1980), pp 111-20
- Manton K.G., Stallard E., Vaupel J.W. Alternative models for the heterogeneity of mortality risks among the aged. *Journal of the American Statistical Association*, vol 81 (1986), pp 635-44
- Manton K.G., Stallard E. *Chronic disease modelling: measurement and evaluation of the risks of chronic disease processes*. Charles Griffin & Company Ltd, London, 1988
- Mode G.J. *Stochastic processes in demography and their computer implementation*. Springer Verlag, Berlin (etc.), 1985
- Moeschberger M.L., David H.A. Life tests under competing causes of failure and the theory of competing risks. *Biometrics*, vol 27 (1971), pp 909-33
- Mulder P.G.H., Hempenius A.L. Analysis of total mortality in the Rotterdam sample of the Kaunas-Rotterdam Intervention Study (KRIS). *Statistics in Medicine*, vol 6 (1987), pp 501-16
- Oakes D. Survival times: aspects of partial likelihood (with discussion). *International Statistical Review*, vol 49 (1981), pp 235-64
- Oakes D. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, vol 84 (1989), pp 487-93
- Prentice R.L., Kalbfleisch J.D., Peterson jr A.V., Flournoy N., Farewell V.T., Breslow N.E. The analysis of failure times in the presence of competing risks. *Biometrics*, vol 34 (1978), pp 541-54
- Prentice R.L., Kalbfleisch J.D. Hazard rate models with covariates. *Biometrics*, vol 35 (1979), pp 25-39
- Schluchter M.D., Jackson K.L. Log-linear analysis of censored survival data with partially observed covariates. *Journal of the American Statistical Association*, vol 84 (1989), pp 42-52
- Slud E., Byar D. How dependent causes of death can make risk factors appear protective. *Biometrics*, vol 44 (1988), pp 265-9
- Stijnen T. *Statistische analyse van levensduurgegevens*. Biomedische toepassingen. Stichting Opleidingen Statistiek, Amsterdam
- Thomas D.C. General relative-risk models for survival time and matched case-control analysis. *Biometrics*, vol 37 (1981), pp 673-86

Tsai S.P., Lee E.S., Hardy R.J. The effect of a reduction in leading causes of death: Potential gains in life expectancy. *American Journal of Public Health*, vol 68 (1978), pp 966-71

Wong O. A competing-risk model based on the life table procedure in epidemiological studies. *International Journal of Epidemiology*, vol 6 (1977), pp 153-9

BIJLAGE A**Berekening van de sterftetabel.****A.1 Inleiding.**

In deze bijlage wordt een aantal van de berekeningen beschreven, waarvan de resultaten in het rapport gepresenteerd zijn. De berekeningen zijn uitgevoerd met een computerprogramma.

De uitgevoerde berekeningen kunnen als volgt onderverdeeld worden:

- Het onder één noemer brengen van de invoergegevens.
- Het bepalen van sterftetabellen op basis van sterftefracties.
- Het bepalen van het belang van doodsoorzaken.
- Het bepalen van sterftetabellen na eliminatie van een doodsoorzaak.

De CBS-gegevens, waarvan bij de berekeningen gebruik gemaakt is, zijn te verdelen in: bevolkingsaantallen, sterfte-aantallen en doodsoorzaak-specifieke sterftefracties, alle per geslacht en leeftijdsklasse voor 1986 en 1987.

A.2 Het onder één noemer brengen van de gegevens.

De gebruikte en gepresenteerde cijfers verschillen op de volgende twee punten van vorm:

- De leeftijdsverdelingen, waarop de cijfers betrekking hebben.
- Absolute aantallen of (relatieve) fracties.

De toegepaste leeftijdsverdelingen bestaan in principe uit vijfjaars klassen, waarbij de 85+ klasse als een restklasse fungeert. De uitzondering hierop is, dat de klasse met leeftijd 0-4 opgedeeld is in:

- drie klassen (0, 1, 2-4) voor de bevolkingsaantallen en totale sterfte-aantallen,
- twee klassen (0,1-4) voor de doodsoorzaak-specifieke sterftefracties.

In de gepresenteerde sterftetabellen is de klasse 0-4 niet opgedeeld.

De totale sterftecijfers zijn gegeven in de vorm van absolute aantallen, de doodsoorzaak-specifieke sterftecijfers in de vorm van sterfte-aantallen per 100.000 van de betreffende leeftijdsklasse.

Omdat bij de berekening van de sterftetabellen gebruik wordt gemaakt van sterftefracties, zijn de genoemde sterftecijfers als volgt bewerkt.

Totale sterfte.

De totale sterftefractie voor iedere vijfjaars leeftijdsklasse (i,i+4) vormt het quotiënt van de totale sterfte en de 'totale' bevolking. De totale sterfte bestaat uit de som van de sterfte-aantallen voor beide jaren. De sterfteaantallen in ieder jaar bestaan uit het produkt van de gegeven sterftefracties en de bevolkingsomvang.

$$\text{sterfte}(i,i+4) =$$

$$\Sigma(jr=1986,1987) \text{ sterftefractie}(jr,i,i+4) * \text{bevolking}(jr,i,i+4)$$

Voor de leeftijdsklasse (0,4) moeten de berekende totale sterfte-aantallen voor de klassen 0, 1, 2-4 gesommeerd worden.

De totale bevolking bestaat uit de som van die voor 1986 en 1987.

$$\text{bevolking}(i,i+4) =$$

$$\Sigma(jr=1986,1987) \text{ bevolking}(jr,i,i+4)$$

De uiteindelijke sterftefractie vormt het (geschaalde) quotiënt van totale sterfte-aantallen en totale bevolking..

$$\text{sterftefractie}(i,i+4) =$$

$$\text{sterfte}(i,i+4) / \text{bevolking}(i,i+4) / 100.000$$

Doodsoorzaak-specifieke sterfte.

De doodsoorzaak-specifieke sterftefractie vormt het quotiënt van de doodsoorzaak-specifieke sterfte en de totale bevolking. De doodsoorzaak-specifieke sterfte bestaat uit de som van de sterftes voor beide jaren.

$$\text{specifieke sterfte}(i,i+4) =$$

$$\Sigma(jr=1986,1987) \text{ specifieke sterfte}(jr,i,i+4)$$

De uiteindelijke doodsoorzaak-specifieke sterftefractie vormt het (geschaalde) quotiënt van specifieke sterfte en totale bevolking.

$$\text{specifieke sterftefractie}(i,i+4) =$$

$$\text{specifieke sterftefractie}(i,i+4) / \text{bevolking}(i,i+4) / 1.000$$

A.3 De berekening van de sterftetabel.

De gepresenteerde sterftetabellen beschrijven de overleving van een fictief geboortecohort met gebruik van actuele sterftecijfers. Deze sterftecijfers vormen gemiddelden over de jaren 1986 en 1987 (zie vorige paragraaf). De sterftecijfers kunnen betrekking hebben op de totale sterfte of op doodsoorzaak-specifieke sterfte.

De gepresenteerde sterftetabellen bevatten de volgende grootheden, waarbij de verkorte notatie tussen haakjes is gegeven: de gemeten sterftefractie (m), die gelijk is aan de gemeten sterftefractie, de overlevingsaantallen (l), en de resterende levensverwachting (e). Daarnaast kunnen de volgende andere grootheden onderscheiden worden, die ook vaak in sterftetabellen gepresenteerd worden: de sterfteaantallen (d), de overlevingsjaren per klasse (L), en de resterende overlevingsjaren (T). Elk van de grootheden wordt voor iedere leeftijdsklasse gegeven. Toegepast zijn vijfjaars leeftijdsklassen, waarbij de 85+ klasse een restklasse vormt. De sterftefracties zijn gegeven, de overige grootheden worden hieruit afgeleid.

De overlevingsaantallen.

Voor ieder beginjaar van een leeftijdsklasse wordt berekend, wat het overlevingsaantal is onder de gegeven sterftefracties. Het beginaantal wordt op (zeg) 100.000 gesteld.

$$l(0) = 100.000$$

$$l(i+5) = l(i) * (1 - m(i, i+4))^5$$

De sterfte-aantallen.

De sterfte-aantallen vormen het verschil tussen de overlevingsaantallen aan het begin van twee opeenvolgende klassen.

$$d(i, i+4) = l(i) - l(i+5)$$

$$d(85+) = l(85)$$

De overlevingsjaren per klasse.

De overlevingsjaren per klasse is het totaal aantal jaren, dat door de individuen van het cohort in een klasse doorgebracht wordt. Indien aangenomen wordt, dat de sterfte regelmatig over de tijd doorgebracht in een klasse verdeeld is, wordt dit aantal jaren slechts bepaald door het gemiddelde overlevingsaantal en de breedte van iedere klasse.

$$L(i, i+4) = (l(i) + l(i+5)) / 2 * 5$$

Voor de 'oudste' klasse is deze rekenwijze niet toepasbaar, omdat de breedte van deze klasse niet gedefinieerd is. Voor deze klasse wordt verondersteld, dat de betreffende sterftefractie voor alle jaren in deze klasse constant is. Dit heeft tot gevolg, dat de overlevingsaantallen een meetkundige reeks vormen. Bovendien wordt de sterfte gedurende een jaar gecentreerd verondersteld. Het in deze klasse doorgebrachte aantal jaren is vervolgens eenvoudig te berekenen.

$$L(85+) = l(85+) * (0.5 + \frac{1-m(85+)}{m(85+)})$$

De resterende overlevingsjaren.

De resterende overlevingsjaren worden terugwaarts berekend uit de overlevingsjaren per klasse.

$$T(85+) = L(85+)$$

$$T(i,i+4) = L(i,i+4) + T(i+5,i+9)$$

De resterende levensverwachting.

De resterende levensverwachting kan worden geschreven als het quotiënt van de resterende overlevingsjaren en de overlevingsaantallen. Deze resterende levensverwachting geldt dus voor de beginleeftijd van iedere klasse en vormt dus niet de gemiddelde levensverwachting over de klasse.

$$e(i) = T(i,i+4) / l(i)$$

A.4 Totale sterfte en verloren levensjaren.

Het belang van de drie genoemde doodsoorzaken wordt gemeten aan de hand van twee criteria: de totale sterfte aan een doodsoorzaak en het aantal verloren levensjaren ten gevolge van de sterfte. Bij de verschillende berekeningen wordt gebruik gemaakt van de grootheden, die in de twee voorgaande paragrafen gedefinieerd en berekend zijn.

Totale sterfte.

De totale sterfte, gemiddeld over de jaren 1986 en 1988, wordt berekend als de som van de sterftes voor de verschillende klassen. De sterfte voor iedere klasse vormt het product van de betreffende (specifieke) sterftekans m en de bevolkingsomvang. In het geval van specifieke sterftekansen worden deze gelijk gesteld aan de gemeten specifieke sterftefracties.

$$\text{tot sterfte} = \sum(i) m(i,i+4) * \text{bevolking}(i,i+4)$$

Verloren levensjaren.

Het aantal verloren levensjaren opzichte van de resterende levensverwachting, eventueel ten gevolge van een specifieke doodsoorzaak, wordt berekend als de som van de verloren levensjaren voor iedere klasse. Het aantal verloren levensjaren voor iedere klasse vormt het product van de sterfte en de gemiddelde resterende levensverwachting. Voor de sterfte kan doodsoorzaak-specifieke sterfte gekozen worden, de resterende levensverwachting heeft betrekking op de totale sterfte.

$$\text{verl levensj} = \sum(i) m(i,i+4) * \text{bevolking}(i,i+4) * \\ (e(i)+e(i+5)) / 2$$

Voor de 85+ klasse wordt de gemiddelde resterende levensverwachting gesteld op de helft van de levensverwachting op leeftijd 85 jaar.

A.5 De sterfte na eliminatie van een doodsoorzaak.

De sterfte na eliminatie van een doodsoorzaak wordt op dezelfde wijze berekend als voor de situatie, waarin alle doodsoorzaken aanwezig zijn. Deze berekeningen zijn beschreven in paragraaf A.3. Het verschil betreft de toegepaste sterftekanzen. In plaats van de oorspronkelijke sterftekanzen (m_{oud}) wordt de sterftekanzen (m_{nieuw}) genomen na eliminatie van een doodsoorzaak met gemeten sterftefractie m_{spec} . De nieuwe sterftekanzen zijn:

In het geval van afhankelijke kansen:

$$m_{nieuw}(i,i+4) = m_{oud}(i,i-4) - m_{spec}(i,i-4)$$

In het geval van onafhankelijke kansen:

$$m_{nieuw}(i,i+4) = 1 - (1 - m_{oud}(i,i+4)) \cdot \frac{1 - m_{spec}(i,i+4)}{m_{oud}(i,i+4)}$$

Bij de berekeningen is steeds gekozen voor afhankelijke kansen.

WEGRIJPELIJST**A**

aannemelijkheidsfunctie de kansverdelingsfunctie van de overlijdenstijdstippen van de individuen van een cohort (c.q. sterfteaantallen per gegeven leeftijdsinterval) als functie van de nog onbekende parameter(s)

partiële - het deel van de volledige aannemelijkheidsfunctie, dat betrekking heeft op de selectie van overledenen gegeven de overlijdenstijdstippen

accelerated-life risico gemodelleerd als schaling van de tijdsas

actuariële model het sterfteproces gemodelleerd door het voor een reeks opeenvolgende leeftijdsintervallen toepassen van de proportional hazards aanname

C

censurering het niet kunnen bepalen van het overlijdenstijdstip van een individu vanwege voortijdige beëindiging van de waarneming

type I - censurering ten gevolge van beperkte waarnemingstijd

type II - censurering ten gevolge van beperkt aantal waarnemingen

cohort groep van personen in de tijd gevolgd

geboorte - groep van personen vanaf de geboorte gevolgd

D

doodsoorzaken

concurrerende - (algemeen) slechts één doodsoorzaak leidt tot sterfte (specifiek) onderling afhankelijke doodsoorzaken

vervangende - onderling onafhankelijke doodsoorzaken

H

heterogeniteit onderscheid tussen de sterfteprocessen van de individuen van een cohort

I

identificeerbaarheid de mogelijkheid om het sterfteproces eenduidig vast te leggen op basis van gemeten overlijdenstijdstippen

K

kansgenererende functie transformatie van de kansverdelingsfunctie van een stochast, met behulp waarvan de momenten van de stochast bepaald kunnen worden

M

maximum likelihood statistische techniek om schatter(s) van nog onbekende parameter(s) van een kansverdelingsfunctie te bepalen

moment karakterisering van de kansverdelingsfunctie van een stochast

 eerste - de verwachting van de stochast

 tweede centrale - de variantie van de stochast

mortaliteitsdruk de (voorwaardelijke) verandering van de overlevingskans

O

overlevingskans (1) kans om een bepaalde leeftijd te bereiken
(2) kans om een jaar ouder te worden

overlijdenstijdstip het tijdstip van overlijden van een individu

 virtueel - het (mogelijk onbekende) tijdstip van overlijden aan een specifieke doodsoorzaak

P

proportional hazards mortaliteitsdrukken, die in verhouding tot elkaar constant zijn

1

risico **accumulatie van mortaliteitsdruk**

multiplicatief - risico gemodelleerd als multiplicatieve factor van referentie-mortaliteitsdruk

3

selectie het (in verwachting) eerder overlijden van
 individuen met een hoog risico

stationariteit de leeftijds specifieke bevolkingsaantallen
veranderen

sterfteproces het stochastische optreden van sterfte (overlijdenstijdstip) van een individu

sterftegraad	voorwaardelijke sterftefractie op bepaalde leeftijd
---------------------	--

sterftekans	voorwaardelijke sterftefractie binnen een jaar
--------------------	---

sterftetabel beschrijft sterftes aantallen afhankelijk van
 de leeftijd

cohort - sterftetabel voor in de tijd gevolgd cohort

complete - discrete sterftetabel, waarbij de leeftijd
in afzonderlijke jaren verdeeld is

continue - sterftetabel, waarbij de leeftijd continu verdeeld is

current - momentane sterftetabel van (een deel van) de bevolking

discrete - sterftetabel, waarbij de leeftijd discreet verdeeld is

ingekorte - discrete sterftetabel, waarbij de leeftijd
in meerjaars-groepen verdeeld is

subject-interval deel van de observatietijd van een individu