

RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEU
BILTHOVEN

Rapport nr. 266750002

**Berekening van levensverwachting uit
mortaliteits follow-up studies**

RT Hoogenveen, AJ Schuit, TLS Visscher,
EJM Feskens, NJD Nagelkerke

juni 1998

Dit onderzoek werd verricht in opdracht en ten laste van de Directie RIVM
in het kader van het project 'Modellering van chronische ziekten en determinanten'

Rijksinstituut voor Volksgezondheid en Milieu (RIVM), Postbus 1, 3720 BA Bilthoven,
telefoon: 030 2749111, fax: 030 2742971

VERZENDLIJST

- 1 Directeur-Generaal RIVM
- 2 Directeur-Generaal van de Volksgezondheid
- 3 Hoofdinspecteur voor de Gezondheidszorg
- 4 Prof dr ir D. Kromhout, directeur Sector Volksgezondheidsonderzoek
- 5 Dr ir J.C. Seidell, hoofd Centrum voor Chronische Ziekten en Milieu-Epidemiologie
- 6 Mw. ir M. L. L. van Genugten
- 7 Drs A. E. M. de Hollander
- 8 Mw. P. van den Hoogen
- 9 Mw. S. Houterman
- 10 Ir J. Jansen
- 11 Mw. ir I. A. M. Maas
- 12 Mw. dr ir H. A. Smit
- 13 Mw. dr ir M. W. W. Verschuren
- 14-18 Auteurs
- 19 Bureau Voorlichting en Public Relations RIVM
- 20 Bibliotheek RIVM
- 21 Bureau Rapportenregistratie RIVM
- 22-50 Bureau Rapportenbeheer RIVM

INHOUDSOPGAVE

Verzendlijst	3
Voorwoord	7
Summary	9
Samenvatting	11
1. Inleiding	13
2. De hazard-functie, overlevingsfunctie en levensverwachting	15
3. Parameterschatting	15
4. Betrouwbaarheidsintervallen voor de levensverwachting	16
5. Kalendertijd en leeftijd	16
6. Benodigde gegevens en implementatie	16
7. Voorbeeldberekening voor Body Mass Index	18
7.1 Inleiding en vraagstelling	18
7.2 Gebruikte gegevens	18
7.3 Analyses	22
7.4 Resultaten	22
7.5 Discussie	23
8. Voorbeeldberekening voor lichamelijke activiteit	24
8.1 Inleiding en vraagstelling	24
8.2 Gebruikte gegevens en analyses	24
8.3 Resultaten	28
8.4 Discussie	28
9. Afsluitende opmerkingen	30
Literatuur	31
Bijlage A. De Gompertz (exponentiële) hazard-functie	33
Bijlage B. Parameterschatting volgens maximum likelihood methode	34
Bijlage C. Betrouwbaarheidsintervallen voor de levensverwachting	35
Bijlage D. Parameter-gevoeligheid van de levensverwachting	36

VOORWOORD

De aanleiding van het onderzoek, waarover in dit rapport verslag gedaan wordt, was een vraag naar statistische methoden en programmatuur om verschillen in overleving te berekenen tussen verschillende klassen voor gegeven risicofactoren (determinanten). Hoewel de levensverwachting een belangrijke maat voor overleving is, worden in wetenschappelijke artikelen over het algemeen slechts mortaliteitsverschillen gepresenteerd in de vorm van hazard ratio's of relative risico's. Omdat SAS (1990) de standaard statistische programmatuur is voor het volksgezondheidsonderzoek op het RIVM, is eerst onderzocht welke mogelijkheden dit pakket biedt om levensverwachtingen te berekenen rekening houdend met links-truncering. Toen bleek dat dit niet mogelijk was, is besloten om deze analyse te implementeren in Splus (StatSci, 1993). Splus biedt namelijk meer mogelijkheden dan SAS om andere dan bestaande statistische analyses te implementeren en uit te voeren. De genoemde analyse is toegepast voor twee voorbeeld-determinanten, waarbij gebruik gemaakt is van gegevens uit twee bekende longitudinale studies: de Zeven-Landen Studie voor Body Mass Index, en de Zutphen-Ouderen Studie voor lichamelijke activiteit.

SUMMARY

Survival analyses of longitudinal studies make often use of the Cox proportional hazards model. Mortality differences between risk groups are presented in terms of estimated regression coefficients or relative risks. The research question we have addressed here relates to the presentation of mortality differences in terms of life expectancy differences. The Cox model is extended to a fully parameterized hazard model using an exponential baseline hazard function. This hazard function is used to calculate the survival function and the rest life expectancy. Two example applications for calculating life expectancy differences are presented. One example concerns differences between Body Mass Index classes using data from the Seven-Countries study. The other example deals with differences between physical activity classes using data from the Zutphen Elderly Study. Attention is paid to the aspect of modelling period (calendar time) and age effects and to the related aspect of left truncation.

The following results have been found:

- Survival curves and life expectancy figures can give more insight in mortality differences in the course of time.
- The data needed are the same as those that are applied in Cox survival analysis. Data on 'the age at the start of the study' is necessary to correct for left truncation.
- In case of too small follow-up times the calculated life expectancy are rather unreliable even though the age range being covered is large.

SAMENVATTING

Bij mortaliteitsanalyses op basis van longitudinale studies wordt veelal gebruik gemaakt van het Cox proportional hazards model. Sterfteverschillen tussen verschillende risicogroepen worden weergegeven in de vorm van geschatte regressiecoëfficiënten of relatieve risico's. De vraag is hoe mortaliteitsverschillen eruitzien voor een andere belangrijke maat als de levensverwachting. Beschreven wordt hoe uitgaande van het Cox model een volledig geparameteriseerde hazard functie afgeleid kan worden, die gebruikt wordt om de overlevingsfunctie en vervolgens de levensverwachting te schatten. Verschillen in levensverwachting worden berekend voor twee voorbeeld-studies: tussen klassen van Body Mass Index op basis van gegevens uit de Zeven-Landen studie, en tussen klassen van lichamelijke activiteit op basis van gegevens uit de Zutphen-Ouderen Studie. Daarbij wordt tevens aandacht geschonken aan de modellering van de kalendertijd en leeftijd, en daarmee samenhangend aan 'links-afkapping' ('left truncation').

Verschillende resultaten zijn gevonden:

- Overlevingsfuncties en levensverwachtingen kunnen meer inzicht geven in de sterfteverschillen over de tijd.
- De benodigde data zijn dezelfde als die bij de standaard Cox analyses worden gebruikt. Om te kunnen corrigeren voor links-truncering dient de 'leeftijd bij aanvang van de studie' bekend te zijn.
- Voor te korte follow-up periodes lijken de berekende levensverwachtingen nogal onbetrouwbaar, ook al wordt een voldoende brede leeftijdsrange bestreken.

1. INLEIDING

Veel epidemiologisch onderzoek vindt plaats in de vorm van longitudinale en follow-up studies. Deze studies worden gekenmerkt door de registratie van verschillende persoonskenmerken op verschillende tijdstippen én, zij het soms in een later stadium, een mortaliteits- en vaak ook morbiditeits-follow-up. Voorbeelden zijn de Zutphen (Ouderen) Studie (Feskens et al., 1993) als onderdeel van de FINE-studie (Menotti A, 1996) en de 'Zeven-Landen' studie (Keys, 1980) en het ConsultatieBureau project in combinatie met een follow-up uitgevoerd in samenwerking met het CBS (Verschuren, 1995).

Meestal vindt de rapportage over de sterfteverschillen tussen verschillende groepen (gedefinieerd middels risicofactor-klassen) plaats aan de hand van relatieve risico's of regressiecoëfficiënten voor gegeven statistische sterftemodellen. Voorbeelden van deze modellen zijn het logistische regressiemodel en het Cox proportional hazards model. Een zinvolle, andere vorm van rapportage is een vergelijking in termen van (verschillen in) overleving en daarmee samenhangend de resterende levensverwachting. De berekening daarvan vereist extra rekenstappen (zie bijv. Mode, 1985). Namelijk, uit de hazard-functie wordt de overlevingsfunctie berekend en daaruit weer de resterende levensverwachting. Deze rekenstappen worden in dit rapport verder uitgewerkt, en resultaten worden gepresenteerd voor twee voorbeelden.

Daarbij wordt kort aandacht geschonken aan het verschijnsel van 'links-afkapping' ('left truncation'): individuen worden bij de follow-up pas vanaf latere leeftijd geobserveerd, zodat sprake is van selectie op overleving. Bij de gebruikelijke vormen van sterfte-analyse, waarbij de kalendertijd als tijdparameter wordt gekozen en de observatie voor alle personen (vrijwel) gelijktijdig begint, is dat geen probleem. Wordt de leeftijd als tijdparameter gekozen, is dit wel een probleem. SAS (1990), de standaard statistische programmatuur in de sector Volksgezondheidsonderzoek, biedt geen mogelijkheid om voor links-truncering te corrigeren, Splus (StatSci, 1993) daarentegen wel.

In hoofdstuk 2 wordt de gebruikte hazard-functie geïntroduceerd en beschreven hoe hieruit de overlevingsfunctie en resterende levensverwachting berekend worden. In hoofdstuk 3 wordt de (maximum likelihood) methode van parameterschatting gepresenteerd. In hoofdstuk 4 wordt beschreven hoe asymptotische betrouwbaarheidsintervallen rondom de geschatte levensverwachting berekend kunnen worden. In hoofdstuk 5 wordt dieper ingegaan op de modellering van kalendertijd en leeftijd. Hoofdstuk 6 bevat een korte toelichting op de modelimplementatie. In hoofdstuk 7 wordt de eerste voorbeeld-berekening gepresenteerd: verschillen in overleving en levensverwachting tussen Body Mass Index klassen gebruik makend van gegevens uit de Zeven-Landen Studie. De tweede voorbeeld-berekening, beschreven in hoofdstuk 8, betreft een toepassing voor lichamelijke activiteit. In hoofdstuk 9 worden, mede op basis van de twee voorbeeld-berekeningen, verschillende conclusies getrokken. Het rapport wordt afgesloten met een literatuuroverzicht en verschillende bijlagen. Daarin komen aan de orde: de Gompertz hazard-functie, de maximum likelihood schatting van de modelparameters, de berekening van de betrouwbaarheidsintervallen voor de levensverwachting, en een gevoeligheidsanalyse.

2. DE HAZARD-FUNCTIE, OVERLEVINGSFUNCTIE EN LEVENSVERWACHTING

De (resterende) levensverwachting op een gegeven leeftijd wordt berekend met behulp van een sterftetafel ('life table'). Deze maakt gebruik van de hazards (sterftekansen) vanaf deze leeftijd. Follow-up studies met een beperkte duur leveren echter slechts sterftekansen op tot een bepaalde leeftijd. Om toch de resterende levensverwachting te kunnen berekenen veronderstellen wij een specifieke vorm van de hazard-functie en dat deze geëxtrapoleerd mag worden naar hogere leeftijden. In het Cox-model wordt de vorm van de baseline hazard functie vrijgelaten. Wij veronderstellen dat deze exponentieel met de leeftijd toeneemt. Een dergelijke exponentieel met de tijd toenemende hazard-functie wordt ook wel aangeduid als Gompertz-functie, en geeft over het algemeen een goede beschrijving van sterftekansen voor hogere leeftijden (Mode, 1985). De zo geparmetriseerde hazard-functie wordt gefit op de data voor een beperkte leeftijdsrange. Het gebruik ervan om de resterende levensverwachting te berekenen veronderstelt dat deze ook geldt voor hogere leeftijden (extrapolatie). Formules voor de Gompertz hazard-functie en daaruit volgende overlevingsfunctie worden gepresenteerd in bijlage A.

De resterende levensverwachting wordt berekend uit de overlevingsfunctie. Deze is namelijk gelijk aan de oppervlakte onder de gestandaardiseerde overlevingsfunctie vanaf de leeftijd waarvoor de resterende levensverwachting berekend wordt. De gebruikte formules zijn gepresenteerd in bijlage A.

3. PARAMETERSCHATTING

In geval van Cox proportional hazards analyse worden de regressie-parameters van de hazard-ratio term geschat met gebruik van een partiële aannemelijkheidsfunctie ('partial likelihood') (Hoogenvveen en Jager, 1990). Deze vorm van likelihood is gedefinieerd conditioneel op de sterftetijdstippen. Maximalisatie van de 'partial likelihood' leidt tot schattingen voor de parameters van het regressiedeel van de hazard-functie, niet die van de (Gompertz) baseline hazard functie. Omdat onze hazard-functie in tegenstelling tot die van het Cox proportional hazards model volledig geparmetriseerd is, kunnen we ook de volledige aannemelijkheidsfunctie ('full likelihood') gebruiken in plaats van de 'partial likelihood'. De 'full likelihood' beschrijft de aannemelijkheid van de parameterwaarden gegeven de data. Maximalisatie van de likelihood-functie leidt aldus tot de meest aannemelijke (maximum likelihood, ML) waarden voor alle modelparameters.

Bij de constructie van de likelihood wordt rekening gehouden met 'links-afkapping' ('left-truncation'): de observaties hoeven niet voor alle individuen vanaf tijdstip 0 begonnen te zijn. De likelihood is uitgeschreven in bijlage B.

4. BETROUWBAARHEIDSINTERVALLEN VOOR DE LEVENSVERWACHTING

Betrouwbaarheidsintervallen voor de resterende levensverwachting kunnen (asymptotisch) worden geschat uit de loglikelihood-functie (logaritme van de likelihood-functie). Vanwege de te verwachten scheefheid naar rechts van het betrouwbaarheidsinterval en de per definitie niet-negativiteit van de levensverwachting wordt eerst een log-transformatie toegepast. De variantie van de resterende levensverwachting hangt af van de gevoeligheid ervan voor variaties in de modelparameters en van de statistische variantie van deze parameters. De varianties kunnen asymptotisch worden geschat uit de loglikelihood-functie (zie bijlage A.3).

5. KALENDERTIJD EN LEEFTIJD

Analoog aan Cox proportional hazards analyse kunnen weer twee ‘tijden’ onderscheiden worden, zowel ‘empirisch’ (leeftijd en kalendertijd) als ‘theoretisch’ binnen het gehanteerde statistische model voor de hazard-functie (als tijdparameter in de baseline hazard-functie, of als covariaat). Dit onderscheid maakt in principe twee modelparametrisaties mogelijk: ‘kalendertijd’ als tijdparameter en ‘leeftijd’ (bij aanvang van de studie) als covariaat, resp. ‘leeftijd’ als tijdparameter en ‘kalendertijd’ (voor gegeven leeftijd) als covariaat. Vanwege de sterke leeftijdsafhankelijkheid van sterfte leidt opname van de covariaat ‘leeftijd’ in een model met de kalendertijd als tijdparameter vrijwel altijd tot een significant betere modelfit (zie bijv. Hoogenveen et al., 1993). Voor opname van de covariaat ‘kalendertijd’ in een model met de leeftijd als tijdparameter is dat zeker niet altijd het geval.

Dit onderscheid tussen beide vormen van modelparametrisaties speelt met name in het algemene geval van Cox proportional hazards analyse, omdat de baseline hazard functie dan vrijgelaten (niet-parametrisch) is. In het geval dat deze exponentieel afhankelijk van de leeftijd verondersteld wordt, zijn de modellen met de leeftijd als tijdparameter en de kalendertijd (voor gegeven leeftijd) als covariaat, óf met de kalendertijd als tijdparameter en de leeftijd (voor gegeven jaar) als covariaat, identiek.

6. BENODIGDE GEGEVENS EN IMPLEMENTATIE

De benodigde gegevens zijn dezelfde als die voor een standaard Cox proportional hazards analyse. Analoog aan de Cox proportional hazards analyse worden gestratificeerde analyses uitgevoerd voor onderscheiden subgroepen binnen de populatie of een ‘gepoolde’ analyse, waarbij voor iedere subgroep (behalve de ‘baseline’ groep) gecorrigeerd wordt middels een extra modelparameter (dummy).

De schattingsmethode en modelanalyses zijn geïmplementeerd in een Splus-computerprogramma (StatSci, 1993) met de volgende onderdelen:

- inlezen ASCII dataset en omzetting in Splus-matrixvorm
- selectie van individuen, namelijk zonder ontbrekende waarnemingen
- in geval van een gegeven continue determinant, constructie van indicator-variabelen voor de klassen gedefinieerd door gegeven klasse-grenzen

In geval van een gestratificeerde analyse worden de volgende stappen voor iedere onderscheiden klasse afzonderlijk uitgevoerd, in geval van een 'gepoolde' analyse voor alle klassen tezamen:

- in geval van gestratificeerde analyse: selectie individuen behorende tot klasse
in geval van niet-gestratificeerde analyse: constructie van indicator-variabelen (dummies)
voor de onderscheiden klassen
- parameterschatting
- schatting van covariantiematrix voor de geschatte parameters
- voor gegeven waarde(n) van de covariaten:
 - plot van survivalfunctie
 - berekening van resterende levensverwachting met betrouwbaarheidsgrenzen.

7. VOORBEELDBEREKENING VOOR BODY MASS INDEX

7.1 Inleiding en vraagstelling

Twee voorbeeld-berekeningen zijn uitgevoerd om de mogelijkheden en onmogelijkheden na te gaan van het berekenen van overlevingsfuncties en levensverwachtingen en het gebruiken hiervan voor de beschrijving van sterfteverschillen. Gekozen zijn de determinanten (risicofactoren) 'Body Mass Index' en 'lichamelijke activiteit'. Bij elk van de berekeningen is gecorrigeerd voor een zinvolle selectie van covariaten.

De eerste toepassing betreft een doorrekening voor Body Mass Index met gegevens van de Zeven-Landen Studie (Keys, 1980). De vraagstelling luidt: vergelijk de mortaliteit (in termen van overlevingsfunctie en levensverwachting) voor verschillende BMI-klassen, gecorrigeerd voor de confounders leeftijd en roken.

7.2 Gebruikte gegevens

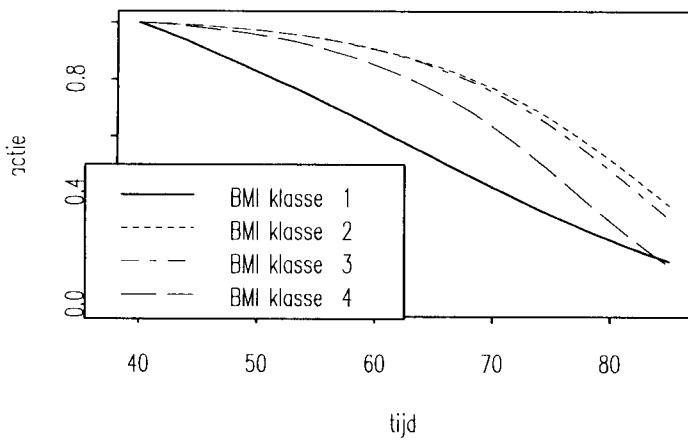
Gebruik is gemaakt van gegevens beschikbaar uit de Zeven-Landen Studie. De volgende variabelen zijn bij de analyse gebruikt:

cohort	cohort nummer
leeftijd	leeftijd bij aanvang van de studie, afgerond naar dichtstbijzijnde verjaardag
BMI	Body Mass Index, klassen onderscheiden op basis van grenzen 18.5, 25 en 30 kg/m ²
roken	rookgedrag, nooit-roken (code 0), ex-roken (code 1-3), roken (code 4-8)
survcode	codering van doodsoorzaak
survtime	waarnemingsduur gemeten in maanden vanaf aanvang studie

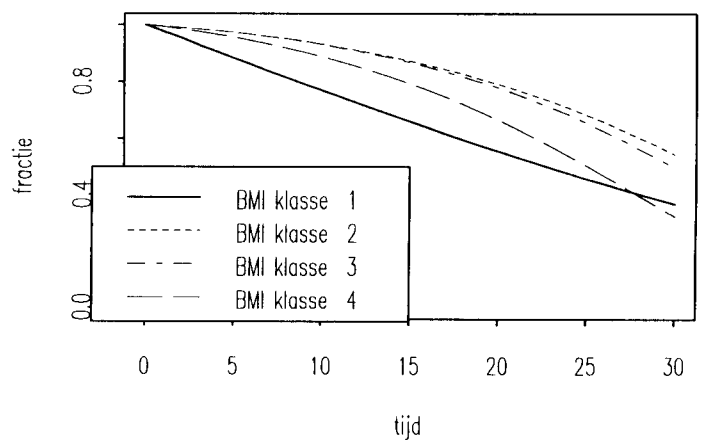
De oorspronkelijke dataset is als volgt bewerkt:

- Waarnemingen voor die individuen zijn meegenomen zonder ontbrekende waarnemingen voor alle bovengenoemde variabelen, met weglating van alle individuen uit de cohorten 13 (Tanushimaru, Japan) en 14 (Ushibuka, Japan).
- Eindtijdstip in jaren, in geval van kalendertijd als tijdparameter: 'survtime'/12, in geval van leeftijd als tijdparameter: 'leeftijd' + 'survtime'/12.
- Bepaling van begintijdstip in jaren, in geval van kalendertijd als tijdparameter: 0, in geval van leeftijd als tijdparameter: 'leeftijd'.
- Event-indicator, 1 indien 'survcode' niet ontbreekt, 0 anders (missing)

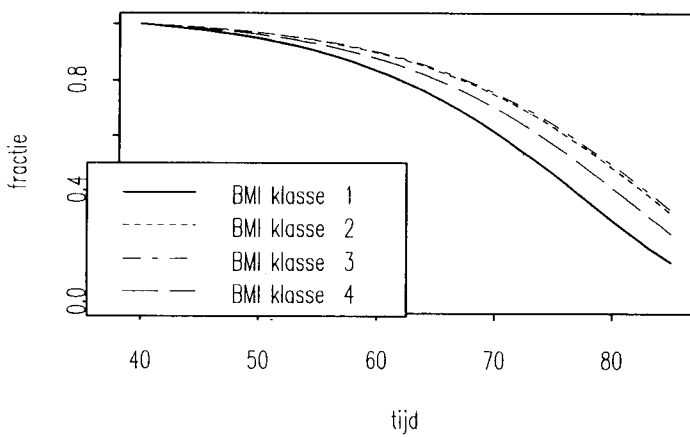
overlevingsfunctie



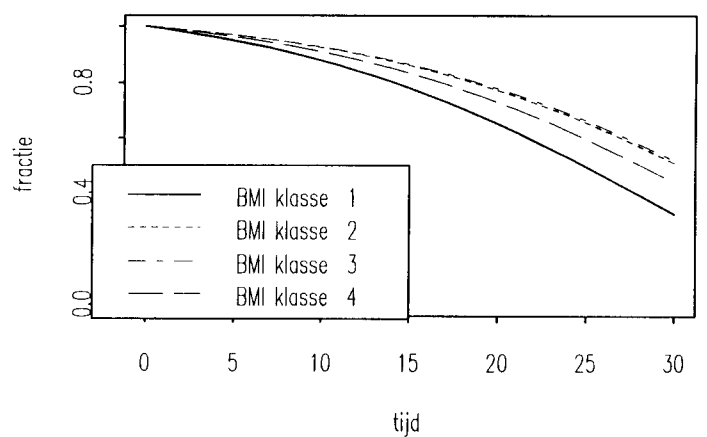
overlevingsfunctie



overlevingsfunctie



overlevingsfunctie



Figuur 1. De overlevingsfuncties voor een niet-roker voor de onderscheiden BMI-klassen op basis van de gestratificeerde analyse en tijdparameter leeftijd (linksboven), tijdparameter kalendertijd (rechtsboven), de gepoolde analyse en tijdparameter leeftijd (linksonder) en tijdparameter kalendertijd (rechtsonder).

Tabel 1: parameterschattingen voor modeltoepassing op Body Mass Index

klasse	α	interc	leeftijd	rook1	rook2	BMI2	BMI3	BMI4	logl
A	gestratificeerde analyse, tijdparameter leeftijd, zie figuur 1								
1	.038±.021	-2.43±2.00		.25±.77	.19±.56				- 576
2	.087±.004	-7.37±0.34		.32±.29	.66±.10				-13898
3	.092±.006	-7.66±0.45		.29±.15	.49±.12				- 7972
4	.093±.012	-7.25±0.96		.06±.26	.04±.23				- 1806
som:									-24252
B	gestratificeerde analyse, tijdparameter kalendertijd, zie figuur 2								
1	.021±.022	-3.81±2.16	.078±.037	.32±.77	.26±.56				- 572
2	.083±.005	-7.65±0.33	.094±.006	.31±.14	.66±.10				-13894
3	.091±.007	-7.74±0.43	.094±.008	.29±.15	.49±.12				- 7972
4	.089±.014	-7.53±0.94	.100±.018	.06±.26	.05±.23				- 1805
som:									-24243
C	'gepoolde' analyse, tijdparameter leeftijd, zie figuur 3								
	.088±.004	-6.79±0.27		.28±.09	.53±.07	-.52±.16	-.56±.17	-.32±.19	-24280
D	'gepoolde' analyse, tijdparameter kalendertijd, zie figuur 4								
	.084±.004	-7.06±0.27	.095±.005	.27±.09	.53±.07	-.51±.16	-.54±.16	-.31±.19	-24274

Noten: Rook1, rook2: variabelen behorende bij de rook-variabelen 'roken' en 'ex-roken'; BMI2, BMI3, BMI4: variabelen behorende bij de opeenvolgende BMI-klassen (BMI1 referentie-klasse); logl: loglikelihood-waarde; voor iedere modelparameter is de geschatte meest aannemelijke waarde gepresenteerd plus tweemaal de standaardfout, berekend uit de Hessiaanse matrix.

Tabel 2: geschatte resterende levensverwachting en verschillen voor Body Mass Index

Resterende levensverwachting:

op leeftijd	40				50				60			
analyse	A	B	C	D	A	B	C	D	A	B	C	D
BMI-klasse												
1	27.2	34.9	32.4	33.5	21.7	24.7	23.9	24.2	17.0	14.6	16.3	15.9
2	38.8	39.1	37.8	38.1	29.8	30.2	28.8	29.2	21.5	21.1	20.6	20.2
3	37.8	37.5	38.1	38.4	28.7	28.8	29.1	29.5	20.4	20.3	20.9	20.4
4	32.9	33.9	35.7	36.4	24.1	24.4	26.9	27.2	16.4	15.8	18.9	18.4

Vershil in resterende levensverwachting met gekozen referentie-klasse BMI-klasse 2

op leeftijd	40				50				60			
analyse	A	B	C	D	A	B	C	D	A	B	C	D
BMI-klasse												
1	-11.6	-4.2	-5.4	-4.6	-8.1	-5.5	-4.9	-5.0	-4.5	-6.5	-4.3	-4.3
3	-1.0	-1.6	+0.3	+0.3	-1.1	-1.4	+0.3	+0.3	-1.1	-0.8	+0.3	+0.2
4	-5.9	-5.2	-2.1	-1.7	-5.7	-5.8	-1.9	-2.0	-5.1	-5.3	-1.7	-1.8

Breedte van de betrouwbaarheidsintervallen rondom de geschatte resterende levensverwachting op leeftijd 50 jaar

analyse	A		B		C		D	
BMI-klasse								
1	2.3	208	2.3	262	19.5	29.2	19.8	29.5
2	23.8	37.4	24.3	37.6	23.7	35.0	24.3	35.1
3	21.2	38.8	21.7	38.2	24.0	35.3	24.5	35.4
4	11.8	49.2	12.2	48.8	21.9	33.0	22.3	33.1

Opmerkingen:

- Voor BMI en de andere covariaten zijn de waarden genomen zoals ze gemeten werden bij aanvang van de studie (1960). Verondersteld is dat deze constant over de tijd zijn.
- In geval van gestratificeerde analyse zijn de aantallen in de opeenvolgende onderscheiden BMI-klassen: 210, 6600, 4000 resp. 799; in geval van de 'gepoolde' analyse 11609.

7.3 Analyses

Verschillende analyses zijn uitgevoerd:

- Gestratificeerd voor elk van de BMI-klassen afzonderlijk en 'gepooled' voor alle BMI-klassen tegelijk. De gebruikte covariaten zijn een constante, de twee rookvariabelen, leeftijd in geval van kalendertijd als tijdparameter, en drie BMI-variabelen in geval van en niet-gestratificeerde analyse.
- Met tijdparameter kalendertijd (en dan extra covariaat leeftijd) of als tijdparameter leeftijd (en dan zonder covariaat kalendertijd) (zie §5).

7.4 Resultaten

7.4.1 Parameterschattingen

In tabel 1 staan de parameterschattingen voor de verschillende uitgevoerde analyses. In figuur 1 zijn de geschatte overlevingsfuncties gepresenteerd voor de onderscheiden BMI-klassen voor een niet-roker. In geval van leeftijd als tijdparameter zijn de overlevingsfuncties gedefinieerd t.o.v. de beginleeftijd (= 40 jaar). In geval van kalendertijd als tijdparameter zijn deze gedefinieerd t.o.v. het begintijdstip van de observatieperiode (1960, model-tijdparameter=0) voor een individu met dan leeftijd 50 jaar. In tabel 2 staan de uit de overlevingsfuncties berekende levensverwachtingen en verschillen daartussen voor de onderscheiden groepen.

7.4.2 Significanties en betrouwbaarheidsintervallen

De betrouwbaarheidsgrenzen rondom de geschatte levensverwachtingen zijn relatief groot. Dit is een gevolg van de sterke gevoeligheid van de resterende levensverwachting voor de parameter α (zie Bijlage B). De 'gepoolde' analyses voor tijdparameter kalendertijd laten een significant verschil zien in de resterende levensverwachting tussen enerzijds de BMI-klasse 1 en anderzijds de klassen 2 en 3.

7.5 Discussie

Vrijwel alle geschatte modelparameters zijn significant van 0 verschillend, behalve relatief veel parameters voor de lichtste BMI-groep en die voor de twee roken-covariaten voor de zwaarste BMI-groep. Het eerste lijkt het gevolg te zijn van het kleine steekproefaantal, het tweede van een inderdaad kleiner effect van roken. De betrouwbaarheidsintervallen rondom de geschatte levensverwachtingen zijn echter breed tot zeer breed. Dit resultaat is niet onverwacht. Het blijkt dat de relatieve onbetrouwbaarheid van de geschatte levensverwachting van dezelfde orde is als die van de geschatte parameter van de exponentiële baseline hazard-functie (zie Bijlage B).

De resultaten laten duidelijke mortaliteitsverschillen zien tussen de onderscheiden BMI-klassen met aan de ene (ongunstige) kant de klassen 1 en 4 en aan de andere (gunstige) kant de klassen 2 en 3. De verschillen in resterende levensverwachting zijn het grootst, indien uitgegaan wordt van de gestratificeerde analyses, maar lijken dan niet-significant. Met opzet wordt hier geschreven 'lijken', omdat het bij de berekeningswijze van de betrouwbaarheidsgrenzen om een eerste orde asymptotische benadering gaat. De verschillen zijn kleiner, maar nog steeds fors en bovendien significant, indien uitgegaan wordt van een 'gepoolde' analyse.

De resterende levensverwachting is duidelijk het kleinst voor de BMI-klasse 1. De resterende levensverwachtingen voor de BMI-klas 2 en 3 verschillen weinig. In geval van een gestratificeerde analyse is die voor klasse 2 het grootst, in geval van een 'gepoolde' analyse die voor klasse 3.

De modelfits en geschatte modelparameters verschillen weinig tussen de modelversies met leeftijd of kalendertijd als tijdparameter. Dit resultaat is in overeenstemming met eerdere analyses uitgevoerd op de Zutphen-studie met het Cox proportional hazards model (Hoogenveen et al., 1993). Voor specifieke doodsoorzaken kunnen de regressiecoëfficiënten echter wel verschillen, maar dat valt buiten onze probleemstelling.

De parameters van de Gompertz baseline hazard-functie verschillen duidelijk voor de onderscheiden BMI-klassen, met name tussen de lichtste klasse enerzijds en de overige klassen anderzijds. Het lijkt erop alsof een zeer lage BMI direct en dus voor relatief lage leeftijden tot verhoogde sterftekansen leidt, terwijl dit voor een hoge BMI pas later geldt en dus voor hogere leeftijden. Dit tijdseffect wordt slechts duidelijk gevonden bij stratificatie.

8. VOORBEELDBEREKENING VOOR LICHAAMELIJKE ACTIVITEIT

8.1 Inleiding en vraagstelling

De tweede toepassing van de beschreven methode om de resterende levensverwachting te berekenen betreft lichamelijke activiteit met gegevens van de Zutphen-Ouderen Studie (Feskens et al., 1993; Bijnen, 1997). De vraagstelling luidt: vergelijk de mortaliteit (in termen van overlevingsfunctie en levensverwachting) voor verschillende categorieën van lichamelijke activiteit, gecorrigeerd voor de confounders leeftijd, roken en alcohol. HDL-cholesterolniveau, diastolische bloeddruk, en Body Mass Index kunnen als intermediaire factoren beschouwd worden. Daarom is hiervoor niet gecorrigeerd.

8.2 Gebruikte gegevens en analyses

Gebruik is gemaakt van de gegevens beschikbaar uit de Zutphen-Ouderen Studie. De volgende variabelen zijn bij de analyse gebruikt:

leeftijd	leeftijd bij aanvang van de studie, afgerond naar dichtstbijzijnde verjaardag
physact	wekelijkse tijdsduur van lichamelijke activiteit (meting 1985), gediscretiseerd tot drie tertielen (Bijnen, 1997)
roken	rookgedrag, wel/niet
alco	alcoholconsumptie, wel/niet
survcode	codering van doodsoorzaak
survtime	waarnemingsduur berekend uit jaar, maand en dag van overlijden, genoteerd in jaar-eenheden
ap, mi, vmi	aanwezigheid van angina pectoris, myocard infarct, verdacht mi bij aanvang van de studie, wel/niet
cva, tia	idem cerebrovascular accident, transient ischemic attack
niddm, iddm	idem (non) insuline afhankelijk diabetes mellitus

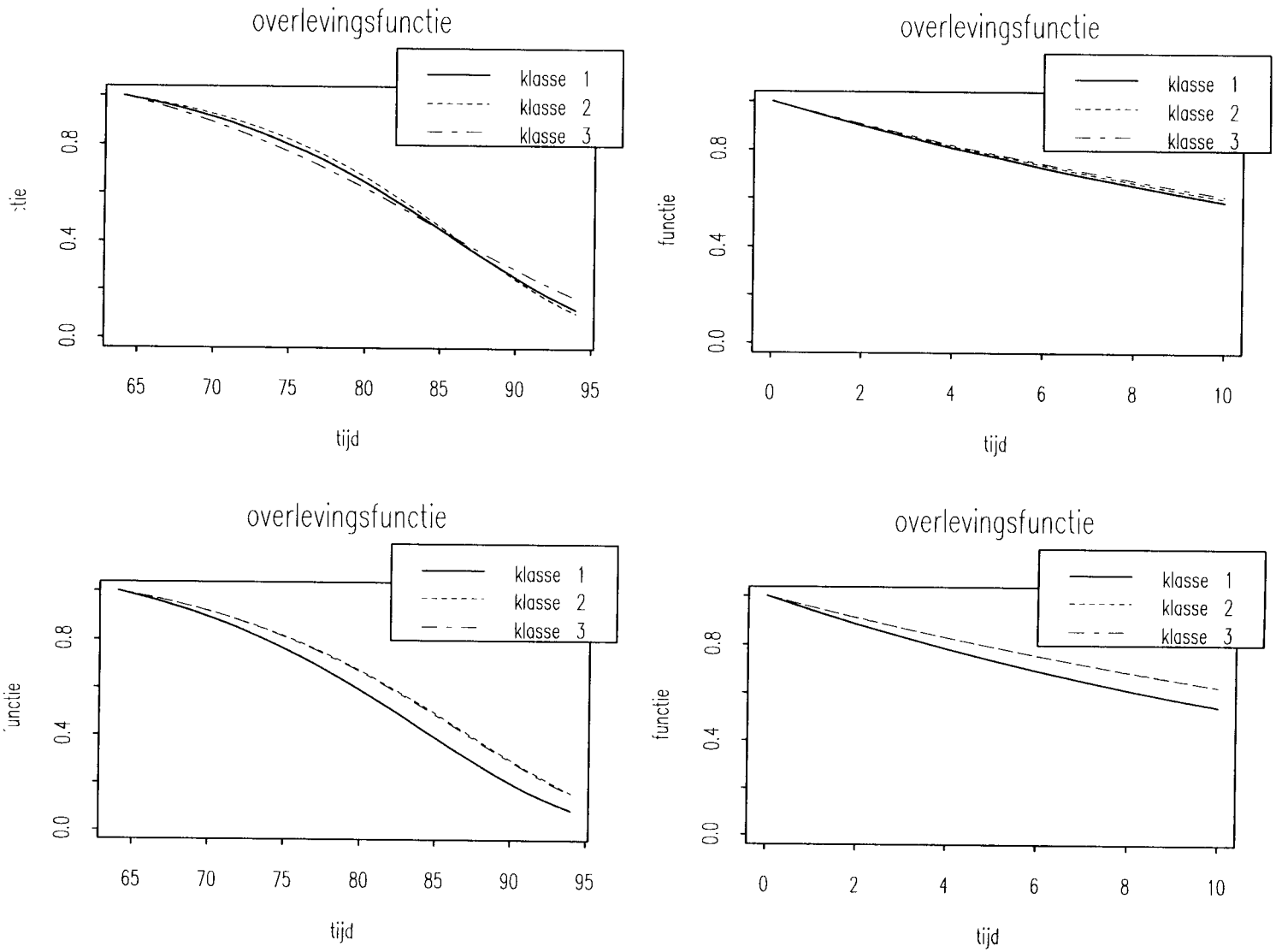
Opmerkingen:

- Roken en alcohol kunnen als confounders beschouwd worden van het effect van lichamelijke activiteit op sterfte.; HDL-cholesterolniveau, diastolische bloeddruk, en Body Mass Index als intermediaire factoren.
- Reeds aanwezige ziekten zijn van invloed op de prognose. Daarom wordt hiervoor steeds gecorrigeerd, waarbij twee ziektegroepen onderscheiden zijn: hart- en vaatziekten (ap, mi, vmi, cva, tia) en diabetes mellitus (niddm, iddm).
- De oorspronkelijke dataset is op analoge wijze bewerkt als voor de eerste voorbeeld-doorrekening (zie §7.2).
- Alle metingen aan lichamelijke activiteit en andere covariaten bij aanvang van de studie (1985).

Tabel 3: parameterschattingen voor voorbeeld-studie op lichamelijke activiteit

klasse	α	interc	leeftijd	rook	alco	act2	act3	logl
A	gestratificeerde analyse, tijdparameter leeftijd, zie figuur 1							
1	.101±.028	-8.63±2.48		.41±.31	4±8e-3			- 598
2	.109±.042	-9.48±3.64		.59±.39	-5±12e-3			- 375
3	.071 (*)	-6.03 (*)		.51±.37	-9±9e-3			- 354
som:								-1327
B	gestratificeerde analyse, tijdparameter kalendertijd, zie figuur 2							
1	.112±.053	-8.63±1.87	.099±.027	.41±.31	4±8e-3			- 598
2	.165±.111	-9.23±2.23	.096±.036	.58±.39	-5±12e-3			- 375
3	.078 (*)	-5.98±2.48	.068±.034	.51±.38	-9±9e-3			- 354
som:								-1327
C	'gepoolde' analyse, tijdparameter leeftijd, zie figuur 3							
	.095 (*)	-8.06±0.27		.50±.18	-1±5e-3	-.26±.19	-.25±.22	-1352
D	'gepoolde' analyse, tijdparameter kalendertijd, zie figuur 4							
	.116 (*)	-8.02±1.10	.091±.014	.50±.18	-1±5e-3	-.27±.21	-.27±.22	-1356

Noten: act2, act3: variabelen behorende bij de opeenvolgende klassen van lichamelijke activiteit (klasse 1 referentie); logl: loglikelihood-waarde; voor iedere modelparameter is de geschatte meest aannemelijke waarde gepresenteerd plus tweemaal de standaardfout, berekend uit de Hessiaanse matrix; *: Splus kon geen standaardfouten berekenen.



Figuur 2. De overlevingsfuncties voor een niet-roker en drinker voor de onderscheiden klassen van lichamelijke activiteit op basis van de gestratificeerde analyse en tijdparameter leeftijd (linksboven), tijdparameter kalendertijd (rechtsboven), de gepoolde analyse en tijdparameter leeftijd (linksonder) en tijdparameter kalendertijd (rechtsonder).

Tabel 4: geschatte resterende levensverwachting en verschillen

Resterende levensverwachting:

op leeftijd		65				75				85			
analyse	act-klasse	A	B	C	D	A	B	C	D	A	B	C	D
1		18.3	17.4	17.5	16.1	11.3	11.1	10.9	10.6	6.2	6.3	6.1	6.3
2		19.0	15.7	19.6	18.0	11.7	10.9	12.6	12.1	6.2	6.9	7.2	7.4
3		19.5	18.7	19.5	17.9	13.3	13.2	12.5	12.1	8.6	8.7	7.2	7.4

Vershil in resterende levensverwachting met gekozen referentie-klasse activiteit-klasse 1

op leeftijd		65				75				85			
analyse	act-klasse	A	B	C	D	A	B	C	D	A	B	C	D
2		0.7	-1.7	2.1	1.9	0.4	-0.2	1.7	1.5	0.0	0.6	1.2	1.2
3		1.2	1.3	2.0	1.8	2.0	2.0	1.6	1.5	2.4	2.4	1.1	1.1

Onder- en bovengrenzen van de betrouwbaarheidsintervallen rondom de geschatte resterende levensverwachting op leeftijd 75 jaar

analyse	A		B		C		D	
act-klasse								
1	11.1	12.2	m	M	7.5	15.6	m	M
2	4.6	28.3	m	M	8.9	17.8	m	M
3	5.1	31.9	m	M	8.8	17.6	m	M

met: m: zeer klein, M: zeer groot; m.a.w.: zeer brede intervallen.

- In geval van de gestratificeerde analyse zijn de aantallen in de opeenvolgende onderscheiden lichamelijke activiteitsklassen: 327, 240 resp. 237; in geval van de 'gepoolde' analyse 804.

Analoge analyses zijn uitgevoerd als bij de eerste voorbeeld-studie (zie Hoofdstuk 7).

8.3 Resultaten

8.3.1 Parameterschattingen

In tabel 3 staan de parameterschattingen voor de verschillende uitgevoerde analyses. In een aantal gevallen (met name analyses voor de laatste activiteitsklasse afzonderlijk) kon het gebruikte Splus-programma geen standaardfouten berekenen voor de parameter α (zie Bijlage B). Dit wijst op brede betrouwbaarheidsintervallen. In figuur 2 zijn de geschatte overlevingsfuncties gepresenteerd voor de onderscheiden klassen van lichamelijke activiteit voor een niet-roker en niet-drinker. In geval van leeftijd als tijdparameter zijn de overlevingsfuncties gedefinieerd t.o.v. de beginleeftijd (= 65 jaar). In geval van kalendertijd als tijdparameter zijn deze gedefinieerd t.o.v. het begintijdstip van de observatieperiode (1985, model-tijdparameter=0) voor een individu met dan leeftijd 75 jaar.

8.3.2 Geschatte levensverwachting plus betrouwbaarheidsintervallen

De geschatte levensverwachtingen en verschillen daartussen zijn gepresenteerd in tabel 4. De bijbehorende betrouwbaarheidsintervallen zijn weer breed tot zeer breed. Zodoende zijn de gevonden verschillen nergens significant.

8.4 Discussie

De berekende verschillen in levensverwachting zijn nergens significant, dus alle vergelijkingen dienen met voorzichtigheid geïnterpreteerd te worden.

De berekende levensverwachtingen zijn aan de hoge kant. Dit lijkt een gevolg van de korte waarnemingsperiode, ook al wordt de leeftijdperiode vanaf 65 jaar goed bestreken.

De resultaten laten mortaliteitsverschillen zien tussen de onderscheiden klassen van lichamelijke activiteit, hoewel de interpretatie ervan lastig is. De verschillen zijn mede afhankelijk van de leeftijd en van de modelparametrisatie. Bij niet-stratificatie lijken matig en intensief bewegen even gunstig ten opzichte van niet-bewegen. Bij stratificatie is matig bewegen pas op hogere leeftijden gunstiger dan niet-bewegen, terwijl intensief bewegen steeds gunstiger wordt met het toenemen van de leeftijd.

De resultaten kunnen vertekend worden door de veronderstelling dat alle covariaten tijdsonafhankelijk zijn. Als mensen bij het ouder worden minder gaan bewegen, zouden de prognoses voor het matig en intensief bewegen gunstiger worden.

De geschatte parameters α van de Gompertz baseline hazard-functie verschillen tussen de onderscheiden activiteitsklassen, zij het niet significant. Conclusies vallen hieruit dan ook niet te trekken.

9. AFSLUITENDE OPMERKINGEN

De weergave van de resultaten van mortaliteitsanalyses in termen van overlevingsfuncties en resterende levensverwachtingen kan duidelijk helpen bij de beschrijving en interpretatie van sterfteverschillen. Met name bij het vergelijken van verschillen tussen verschillende subgroepen over de tijd en over de leeftijd. Bovendien zijn de grafische weergaves van de geschatte overlevingsfuncties mooie hulpmiddelen om artikelen of rapporten te verlevendigen! Wel moeten alle resultaten (zoals gewoonlijk) met enige voorzichtigheid geïnterpreteerd worden.

De levensverwachting wordt berekend over de gehele leeftijdsperiode vanaf een gegeven leeftijdsgrens. Ook al overdekt de leeftijdsrange van de studie deze leeftijdsperiode behoorlijk, toch lijkt een onvoldoende lange follow-up periode tot minder betrouwbare schattingen van de levensverwachting te leiden.

Bij het vergelijken van de resterende levensverwachtingen tussen verschillende subgroepen in termen van gevoeligheid zijn twee verschillende conclusies mogelijk, althans op basis van de door ons uitgevoerde analyses. De resterende levensverwachting is enerzijds een gevoelige indicator, omdat de verschillen tussen de onderscheiden risicofactor-klassen behoorlijk groot kunnen zijn. Anderzijds een ongevoelige indicator, omdat de betrouwbaarheidsintervallen breed zijn vanwege de sterke gevoeligheid ervan voor de te schatten parameter van de Gompertz baseline hazard-functie (zie Bijlage B).

Dit verschil hangt samen met de keuze of men wel of niet stratificeert. Uit de analyses blijkt dat de baseline hazard functies tussen de onderscheiden klassen nogal variëren. Natuurlijk kan de keuze voor stratificatie in statistische zin opgelost worden door modelparametrisaties te vergelijken met behulp van de likelihood-ratio toets. Echter, een grafische weergave van de gestratificeerd geschatte overlevingsfuncties kan ook heel verhelderend werken.

De gehanteerde modelparametrisaties zijn niet geschikt om leeftijds-, periode- en cohort-effecten te onderscheiden. Nader onderzoek dient uit te wijzen of bij veranderingen van risico's over de tijd, zoals bijvoorbeeld gezien in de eerste voorbeeldberekening voor Body Mass Index, van vertragingseffecten over de tijd of leeftijdseffecten sprake is. Dit nadere onderzoek valt buiten de orde van ons eerste terreinverkennde werk (zie bijv. Visscher et al., 1998).

De huidige implementatie van het model in de vorm van een Splus-computerprogramma is voorlopig nog een beperking voor een bredere toepassing ervan, maar wellicht komt in de nabije toekomst een SAS-implementatie beschikbaar.

LITERATUUR

Bijnen FCH. Physical activity and cardiovascular disease risk among elderly men. The Zutphen Elderly Study 1985-1995. Thesis. Universiteit Utrecht, Utrecht, 1997

Cox DR, Oakes D. Analysis of survival data. Chapman&Hall, London etc., 1984

Feskens EJM, Bloemberg BPM, Pijls LTJ, Kromhout D. A longitudinal study on elderly men: the Zutphen Study. In: Schroots JJF, ed. Aging, health and competence. Elsevier Science Publishers BV, 1993

Hoogenveen RT, Feskens EJM, Heisterkamp SH, Lezenne Coulader C de, Bloemberg BPM. Resultaten van een analyse van de Zutphen-studie in het kader van concurrerende doodsoorzaken. Rapport nr. 441111001. RIVM, Bilthoven, 1993

Hoogenveen RT, Jager JC. Overlevingsduuranalyse en concurrerende doodsoorzaken. Een inleiding. Rapport nr. 958606001. RIVM, Bilthoven, 1990

Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. John Wiley & Sons, New York etc., 1980

Keys A. Seven countries: a multivariate analysis of death and coronary heart disease. Harvard University Press, Cambridge (Mass.), 1980

Menotti A, Kromhout D, Nissinen A, Giampaoli S, Seccareccia F, Feskens E, Pekkanen J, Tervahautu M. Short term all-cause mortality and its determinants in elderly populations in Finland, The Netherlands and Italy. The FINE-study. *Prev Med*, vol 25 (1996), pp 319-326

Meijer J, Geunds HA van, Sluijter DP. CB Heart Project in the Netherlands. Screening for risk factors of CHD in consultation bureaus for tuberculosis. *Hart Bulletin*, vol 7 (1976), pp 42-46

Mode CJ. Stochastic processes in demography and their computer implementation. Springer-Verlag, Berlin [etc.], 1985

SAS. Reference Version 6. SAS Institute, Cary, 1990

Statistical Sciences Inc. S-Plus User's Manual, version 3.2. StatSci (a division of MathSof Inc.), Seattle, 1993

Verschuren WMM, Kromhout D. Total cholesterol and mortality at a relatively young age: do men and women differ? *British Medical Journal*, vol 311 (1995), pp 779-783

Visscher TLS, Seidell JC, Feskens EJM, Menotti A, Blackburn H, Kromhout D. Under- and overweight in ageing men in relation to mortality: the Seven Countries Study, 1998. In preparation

BIJLAGE A. DE GOMPERTZ (EXPONENTIËLE) HAZARD-FUNCTIE

De Gompertz (exponentiële) hazard-functie heeft de vorm:

$$h(u) = C \alpha \exp(\alpha u)$$

Daaruit volgt voor de overlevingsfunctie:

$$S(u) = \exp\{ - C [\exp(\alpha u) - 1] \}$$

met: u: tijdparameter, α : modelparameter (regressiecoëfficiënt voor tijdparameter). De resterende levensverwachting op leeftijd t wordt voor willekeurige gegeven overlevingsfunctie als volgt gedefinieerd (uitgaande van een continue tijdparameter):

$$\text{levensverwachting } e(t) = \int_{u \geq t} S(u) du / S(t)$$

met: t: leeftijd waarvan de resterende levensverwachting is gedefinieerd. Voor de levensverwachting bij geboorte wordt gekozen $t=0$. De levensverwachting is gelijk aan de oppervlakte onder de 'conditionele' overlevingsfunctie, dat wil zeggen de overleving conditioneel op overleving tot tijdstip t. Substitutie van de Gompertz hazard-functie leidt tot de volgende formule:

$$\text{levensverwachting } e(t) = \int_{u \geq t} \exp\{ - C [\exp(\alpha u) - \exp(\alpha t)] \} du$$

Aansluitend op het Cox proportional hazards model wordt de hazard-functie afhankelijk verondersteld van covariaten volgens een lineaire regressie-verband. De verschillende genoemde variabelen kunnen dan als volgt uitgewerkt worden:

$$\text{hazard functie } h(u; z) = \exp(\beta' z) \alpha \exp(\alpha u)$$

$$\text{overlevingsfunctie } S(u; z) = \exp\{ - \exp(\beta' z) [\exp(\alpha u) - 1] \}$$

$$\text{levensverwachting } e(t; z) = \int_{u \geq t} \exp\{ - \exp(\beta' z) [\exp(\alpha u) - \exp(\alpha t)] \} du$$

met: z: covariaten (vector-vorm notatie), α, β : regressiecoëfficiënten. Enkele opmerkingen:

- Verondersteld is dat de covariaten tijdsafhankelijk zijn.
- De zo geparametriseerde overlevingsfunctie behoort aldus tot de familie van de Gompertz-functies.
- De term ' $\exp(\beta' z)$ ' kan geïnterpreteerd worden als een hazard ratio; de term ' $\alpha \exp(\alpha u)$ ' beschrijft de baseline hazard-functie.

- Zo geparametriseerd is deze hazard-functie gelijk aan die van het Cox proportional hazards model, met echter een gegeven geparametriseerde in plaats van een vrije niet-parametrische baseline hazard-functie.

BIJLAGE B. PARAMERSCHATTING VOLGENS MAXIMUM LIKELIHOOD METHODE

In geval van Cox proportional hazards analyse worden de regressie-parameters van de hazard-ratio term geschat met gebruik van een partiële aannemelijkheidsfunctie ('partial likelihood'). Maximalisatie van de 'partial likelihood' leidt tot schattingen voor de parameter β van het regressiedeel van de hazard-functie, niet voor de parameter α van de baseline functie. Omdat de Gompertz hazard-functie volledig geparametriseerd is kunnen we de volledige aannemelijkheidsfunctie ('full likelihood') gebruiken. De likelihood-functie beschrijft de aannemelijkheid van de parameterwaarden gegeven de data. Maximalisatie van de likelihood-functie leidt aldus tot de meest aannemelijke (maximum likelihood, ML) parameterschattingen.

Bij de constructie van de likelihood wordt rekening gehouden met links truncering ('left-truncation'): de observaties hoeven niet voor alle individuen vanaf tijdstip 0 begonnen te zijn. De loglikelihood (logarithme van de likelihood-functie) wordt aldus:

$$l(\alpha, \beta; t_{ib}, t_{ic}, M, N) \propto \sum_{i \in N} \log S(t_{ic}; z_i) - \sum_{i \in N} \log S(t_{ib}; z_i) + \sum_{i \in M} \log h(t_{ic}; z_i)$$

met: i : index over individuen, N : totale populatie, t_{ic} : einde van waarnemingsperiode voor i , z_i : covariaten van i , t_{ib} : begin van waarnemingsperiode voor i , M : populatie met event (sterfte) op t_{ic} .

Enkele opmerkingen:

- De afhankelijkheid van de hazard- en survival-functie van de modelparameters α en β is niet expliciet genoteerd.
- Wordt voor alle begintijdstippen gekozen $t_{ib}=0$, dan geldt $S(t_{ib}; z_i)=1$ en $\log(S(t_{ib}; z_i))=0$ en verdwijnt deze term uit de loglikelihood en resteert de klassieke vorm van de loglikelihood-functie.

Substitutie van de gegeven hazard- en survival-functie leidt tot de volgende vorm van de likelihood-functie:

$$l(\alpha, \beta; t_{ib}, t_{ic}, M, N) \propto \sum_{i \in N} \exp(\beta' z_i) \{ \exp(\alpha t_{ib}) - \exp(\alpha t_{ic}) \} + \sum_{i \in M} \beta' z_i + \alpha \#M + \alpha \sum_{i \in M} t_{ic}$$

met: $\#M$: aantal events.

BIJLAGE C. BETROUWBAARHEIDSINTERVALLEN VOOR DE LEVENSVERWACHTING

De betrouwbaarheidsintervallen voor de resterende levensverwachting kunnen worden berekend uit de loglikelihood-functie. Vanwege de te verwachten scheefheid naar rechts van het betrouwbaarheidsinterval en de per definitie niet-negativiteit van de levensverwachting wordt eerst een log-transformatie toegepast. In eerste orde benadering geldt:

$$e(t;z)_{u,i} \approx e(t;z) \exp\{ \pm 2 \sqrt{\text{var}[\log(e(t;z))]} \}$$

$$\text{var}\{\log(e(t;z))\} \approx \sum_{p,q} \delta \log(e(t;z)) / \delta c_p \delta \log(e(t;z)) / \delta c_q \text{cov}_{pq}$$

met: $e(t;z)_{u,i}$: boven- resp. ondergrens van 95% betrouwbaarheidsinterval, var: variantie, δ/δ : partiële eerste orde afgeleide, p,q: indices over de parameters (regressiecoëfficiënten), $c = [\alpha \beta]$: parameters genoteerd in vector-vorm, cov: covariantiematrix van geschatte modelparameters, met $\text{cov}_{pq} = \text{cov}(c_p, c_q)$. De covariantiematrix kan benaderd worden middels de (Hessiaanse matrix van de) loglikelihood:

$$\text{cov} \approx - \{ \text{DD } l(c)_{c=c_0} \}^{-1}$$

met: DD: tweede orde (matrix) afgeleide, c_0 : geschatte, meest aannemelijke (ML) parameterwaarden.

BIJLAGE D. PARAMETER-GEVOELIGHEID VAN DE LEVENSVERWACHTING

De resterende levensverwachting op leeftijd t op basis van het Gompertz-model is gedefinieerd als (met weglating van de afhankelijkheid van de covariaten):

$$e(t) = \int_{u \geq t} \exp\{ -C [\exp(\alpha u) - 1] \} du / \exp\{ -C [\exp(\alpha t) - 1] \} =$$

$$\int_{u \geq t} \exp\{ -C [\exp(\alpha u) - \exp(\alpha(u-t))] \} du =$$

$$\alpha^{-1} \int_{u \geq 0} \exp\{ -C \exp(\alpha t) [\exp(\alpha u) - 1] \} du$$

De logaritmische ervan luidt:

$$\log(e(t)) = -\log(\alpha) + \log\left\{ \int_{u \geq 0} \exp\{ -C \exp(\alpha t) [\exp(\alpha u) - 1] \} du \right\}$$

De asymptotische variantie hiervan kan bepaald worden met behulp van de afgeleide naar de parameters en de covariantie-matrix van de geschatte parameters. De covariantiematrix is asymptotisch gelijk aan de inverse van de Hessiaanse matrix. Dit is de tweede orde matrix-afgeleide van de loglikelihood naar de parameters.

Afgeleide naar parameter α :

$$\delta/\delta\alpha \log(e(t)) = -\alpha^{-1} - Ct \exp(\alpha t) \int_{u \geq 0} \exp\{ -C \exp(\alpha t) [\exp(u) - 1] \} [\exp(u) - 1] du / \{ \alpha e(t) \}$$

$$\text{met: } Ct \exp(\alpha t) \int_{u \geq 0} \exp\{ -C \exp(\alpha t) [\exp(u) - 1] \} [\exp(u) - 1] du = t - Ct \exp(\alpha t) e(t)$$

$$\text{Dus: } \delta/\delta\alpha \log(e(t)) = -\alpha^{-1} \{ 1 + t e(t)^{-1} - Ct \exp(\alpha t) \}$$

Afgeleide naar parameter C :

$$\delta/\delta C \log(e(t)) = -\exp(\alpha t) \int_{u \geq 0} \exp\{ -C \exp(\alpha t) [\exp(u) - 1] \} [\exp(u) - 1] du / \{ \alpha e(t) \}$$

$$\text{met: } \exp(\alpha t) \int_{u \geq 0} \exp\{ -C \exp(\alpha t) [\exp(u) - 1] \} [\exp(u) - 1] du = C^{-1} - \alpha \exp(\alpha t) e(t)$$

$$\text{Dus: } \delta/\delta C \log(e(t)) = -\alpha^{-1} \{ \alpha \exp(\alpha t) - C^{-1} e(t)^{-1} \}$$

Deze resultaten kunnen relatief eenvoudig geïnterpreteerd worden door de resterende levensverwachting in eerste orde te benaderen, d.w.z door de eerste orde benadering van de cumulatieve hazard te nemen:

$$e(t) \approx \int_{u=0}^{\infty} \exp\{-C \exp(\alpha t) u\} du = \alpha^{-1} C^{-1} \exp(-\alpha t)$$

Substitutie in de formules voor de beide afgeleiden leidt tot:

$$\text{mbt } \alpha \quad \{ 1 + t/e(t) - Ct \exp(\alpha t) \} \approx 1 - (1-\alpha) Ct \exp(\alpha t)$$

$$\text{mbt } C \quad \alpha \exp(\alpha t) - C^{-1} e(t)^{-1} \approx 0$$

Substitutie van deze benaderingen in de formule voor de variantie van $\log(e(t))$ leidt tot:

$$\text{var}(\log(e(t))) \approx$$

$$\alpha^{-2} \{ 1 + t/e(t)^{-1} - Ct \exp(\alpha t) \}^2 \text{var}(\alpha) +$$

$$\alpha^{-2} \{ \alpha \exp(\alpha t) - C^{-1} e(t)^{-1} \}^2 \text{var}(C) +$$

$$2 \alpha^{-2} \{ 1 + t/e(t) - Ct \exp(\alpha t) \} \{ \alpha \exp(\alpha t) - C^{-1} e(t)^{-1} \} \text{cov}(\alpha, C)$$

Hieruit blijkt, dat de relatieve onbetrouwbaarheid van de berekende levensverwachting evenredig is met die van de gechatte parameter α .