

TEL-työkyvyttömyyteen liittyvät todennäköisyydet ja niiden mallittaminen

**SHV-harjoitustyö
Mikko Karpoja
8.3.1998**

SISÄLLYS:

J. Johdanto	1
X. Malli	2
X.1 Taustaa	2
X.2 Jäljellä olevan ajan T jakaumaan liittyvää käsitteistöä	2
X.3 Ajan T kuluessa tapahtuvat suoritukset	4
X.4 Satunnaismuuttujan T ja ajanjakson päättävän tapahtuman J yhteisjakauma	7
X.5 Edut, kun päättymissyty vaikuttavat	9
X.6 Mallin varianssit	10
Y. Tilaluokkamallit	11
Y.1. Taustaa	11
Y.2. Tilaluokkamallin johtaminen	12
Y.3. Kaksi approksimaatiota	14
Z. Z-malli	16
Z.1 Taustaa	16
Z.2 Z-mallin todennäköisyydet	16
Z.3 Z-mallin muoto ja sisältö	17
Z.4 Nykyisen Z-mallin sisällön tarkastelua	18
Z.5. Z-mallin pääoma-arvot	21
Z.6 Työkyvyttömyyden ja työkyvyttömyyseläkkeen alkamisen erottaminen	22
Z.7 Z-malli tässä työssä	23
A. Aineiston käsittely	25
A.1 Aineiston epäparametrinen sovittaminen	25
A.2 Aineiston parametrinen sovittaminen	29
A.3 Aineiston käyttö tässä työssä	31
S. Sovitettavat mallit	35
S.1 Jakauman sovittaminen	35
S.2 Suhteellisen vaaran malli	36
S.3 Kiihtyneen tapahtuma-ajan malli	37
S.4. Muita mallittamistapoja	38
S.5. Mallin valintakriteeri	39
T. Tulokset	40
T.1. Kuolevuus ennen sairastumista	40
T.2. Sairastuvuus	41
T.3 Parantuvuus ja kuolevuus työkyvyttömyyseläkkeeltä	43
T.4 Johtopäätöksiä	46

Lähteet

Liitteet Z, T ja S

Abstract:

In this paper the disability insurance model has been developed in the survival analysis context. First the needed disability models for Finnish mandatory pension program (TEL) have been discussed. In present work disability insurance has been developed using multiple state model. The differences of this model and the survival analysis model developed here are shortly discussed. At the moment in TEL the so-called Z-model is being used; the use of the Z-model in this work is briefly discussed. Some examples how survival analysis models can be estimated for disability data are given.

Keywords: disability insurance, survival analysis, accelerated life model

J. Johdanto

Tässä työssä tarkastellaan TEL:n työkyvyttömyysilmiötä todennäköisyyslaskennan näkökulmasta. Tavoitteena on pystyä määrittelemään työkyvyttömyyden kestoon liittyvät todennäköisyyskäsitteet ja antaa esimerkkeinä tapoja, joilla kestoa voidaan mallittaa. Työkyvyttömyyden kestoa on pystyttävä mallittamaan, koska työeläkeyhtiöiden on perittävä maksunsa työkyvyttömyysosa niin, että maksu riittää eläkkeiden maksamiseen. Työkyvyttömyyseläkkeet koostuvat kahdesta osasta ns. yhteisesti kustannettavasta osasta ja rahastoidusta osasta. Yhteisesti kustannettavia eläkkeen osia varten työeläkeyhtiöt perivät vuosittain kollektiivisesti maksua juuri niin paljon, että nämä tuona vuonna maksussa olevat eläkkeen osat voidaan maksaa. Rahastoitu osa nimensä mukaisesti rahastoidaan ennalta joko eläkkeen alkaessa tai työntekijän vielä ollessa työssä, joten on pystyttävä ennustamaan kuinka kauan työkyvyttömyyseläke kestää. Tilanteessa, jossa osa eläkkeestä rahastoidaan ennen varsinaista sairastumista, tarvitaan malli myös työkyvyttömyyden alkamiselle ja ajalle ennen sairastumista.

Tässä harjoitustyössä on TEL:n työkyvyttömyyden kestoa tarkasteltu ns. elinaikamallina. Luvuissa X on ensin esitelty elinaikamalleihin liittyvät peruskäsitteet - siinä on käsitelty yleisesti tietyssä tilassa pysymiseen liittyvää kestoa. Sen jälkeen on johdettu tässä ympäristössä odotusarvot, joita tarvitaan TEL-työkyvyttömyyseläkkeiden rahastoitujen osien laskemiseksi. Luvussa Y on käsitelty viime vuosina on yleisimmin työkyvyttömyyden mallittamiseksi käytettyä ns. tilaluokkamallia ja selvitelty sitä, miten se on yhteydessä luvun X mukaisiin malleihin. Nyt voimassa olevissa laskuperusteissa TEL:n mukaan työkyvyttömyystodennäköisyydet on laskettu ns. Z-mallin avulla. Tätä mallia ja sen käyttöä tässä harjoitustyössä on käsitelty luvussa Z. Elinaikamalleissa oman tarkastelunsa vaatii aineiston käsittely. Yleensä malleja sovitettaessa voidaan saada aineistoksi otos tutkittavasta ilmiöstä. Elinaikamalleihin liittyy kuitenkin omana piirteenään se, että käytettävissä olevana aikana kaikki havainnot eivät pääty tutkittavaan päättymistapahtumaan - osa katoaa seurannasta tai seuranta on pakko päättää ennenkuin kaikki jaksot ovat päättyneet. Tätä ilmiötä ja sen seurauksia on tarkasteltu luvussa A, kuten myös käytössä ollut aineistoa. Luvussa S on esitelty sovitettavat mallit ja luvussa T on kuvattu saadut tulokset sovitetusta malleista.

Tässä harjoitustyössä ei ole käytetty vakuutusmatemaattista merkintätapaa, sen sijaan on käytetty elinaikamallien yhteydessä käytettyjä todennäköisyyslaskennan merkintöjä.

X. Malli

X.1 Taustaa

Tarkastellaan aluksi henkilöä, jonka ikä on x . Henkilö on tietyssä tilassa $Z = Z(x) = a$. Tila on esimerkiksi elossa oleminen, työkykyisenä oleminen tai työttömänä oleminen. Merkitään kirjaimella T aikaa, jonka henkilö viettää samassa tilassa. Iässä $x+T$ henkilö siirtyy toiseen tilaan $Z(x+T) = b$. Ajanjakson *päättävä tapahtuma* J on esimerkissä kuolema, työkyvyttömäksi siirtyminen tai työn saaminen. Ajanjakson voi päättää jokin useammasta eri tapahtumasta. Kullakin päättymisellä voi olla eri seuraus, kuitenkin alkutila a ja päättävä tapahtuma $J=j$ määräävät yksikäsitteisesti uuden tilan b .

Olkoon T , $Z = Z(t)$ ja J satunnaismuuttujia, jotka selvästi riippuvat sekä alkuiästä että alkutilasta. Nämä ehdollistavat suureet on jätetty merkitsemättä, joskin jatkossa on tilanteita, joissa niitä on syytä korostaa, jolloin ne merkitään näkyviin. Koska tämän luvun mallissa henkilön on oltava samassa tilassa koko tarkastelun ajan, emme nyt tarvitse satunnaismuuttujaa Z mukana merkinnöissä ja tarkastelemme lähemmin vain satunnaismuuttujia J ja T , vaikkakin myöhemmin tarvitsemme merkinnän tilanteelle, jossa tila $Z=a$ on päättynyt ja ollaan tarkastelemassa uuden tilan $Z=b$ päättymistä. Ajatellaan aluksi myös, että ajanjakson päättävä tapahtuma J on aina sama tai yhtäpitävästi, että ajanjakson päättävät tapahtumat on yhdistetty, jolloin voimme jättää myös satunnaismuuttujan J merkinnöistä toistaiseksi pois ja näin päädyimme yksinkertaistettuun merkintään *jäljellä olevan ajan* T jakaumalle $P(T \leq t) = F_T(t) = F(t)$, mikä kertoo todennäköisyyden, että henkilön pysyvä olo alkutilassa päättyy viimeistään ajan t kuluttua alkuiästä x . On huomattava, että kaiken aikaa $T > 0$.

Tarkastelu tässä luvussa noudattelee Bowersia ym. [1986].

X.2 Jäljellä olevan ajan T jakaumaan liittyvää käsitteistöä

Oletetaan kaiken aikaa, että on olemassa tiheysfunktio $f_T(t) = f(t)$ s.e. $F(t) = \int_0^t f(t) dt$.

Otetaan oma merkintänsä todennäköisyydelle $P(T > t)$, jolla henkilö viettää alkutilassa vähintään ajan t tilassa, merkitään $S(t) = P(T > t) = 1 - F(t)$ ja kutsutaan sitä *elossaolotodennäköisyydeksi*. Käsite elossaolotodennäköisyys on käytössä elinaikamalleissa. Tulevaa käyttöä varten tarpeellisia ovat eräät ehdolliset todennäköisyydet. Todennäköisyydellä

$$\frac{1 - F(s+t)}{1 - F(s)} = \frac{S(s+t)}{S(s)} = \frac{P(T > s+t)}{P(T > s)} = P(T > s+t \mid T > s) = 1 - F(s+t \mid T > s)$$

henkilö viettää vähintään ajan t tilassa $Z=a$ pysyttyään jo sitä ennen ajan s tilassa $Z=a$.

Jäljellä olevan ajan T odotusarvo saadaan suoraan määritelmästä. Perinteisessä vakuutusmatematiikassa on se kuitenkin totuttu kirjoittamaan määritelmästä poiketen. Tätä sekä myöhempää käyttöä varten täytyy esittää seuraava apulause.

Lemma X.1

Olko T jatkuvasti jakautunut ei-negatiivinen satunnaismuuttuja, jonka kertymäfunktio on $F(t)$ ja tiheysfunktio $f(t)$. Olko $z(t)$ ei-negatiivinen, monotoninen ja derivoituva funktio välillä $[0, \infty)$. Tällöin

$$\begin{aligned} E(z(T)) &= \int_0^{\infty} z(t) \cdot f(t) dt \\ &= z(0) + \int_0^{\infty} z'(t) \cdot (1 - F(t)) dt. \end{aligned}$$

Jäljellä olevan ajan odotusarvoksi saadaan

$$\begin{aligned} E(T) &= \int_0^{\infty} t \cdot f(t) dt \\ &= \int_0^{\infty} (1 - F(t)) dt, \end{aligned}$$

joista edellinen seuraa suoraan määritelmästä ja jälkimmäinen on perinteinen vakuutusmatematiikan esitystapa.

Määritellään seuraavaksi *intensiteetti*. Elinaikamalleissa käsitettä kutsutaan myös ilmaantuvuudeksi ja esimerkiksi henkivakuutusmatematiikassa kuolevuudeksi.

Määritelmä X.1

Olko $\mu_x(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$ päättävyyden *intensiteetti* ajan hetkellä t , jolloin on myös $\mu_x(t) = \frac{f(t)}{1 - F(t)}$

Tässä kohden on tarpeellista korostaa alaindeksillä sitä, että intensiteetti voi riippua tarkastelun alkuhetkestä eli alkuiästä x . On tietysti huomattava, että myös T :n jakauma riippuu näinollen alkuiästä x .

Nyt nähdään, että

$$\mu_x(t) = \frac{f(t)}{1-F(t)} = -\frac{d}{dt} \ln(1-F(t)), \quad \Rightarrow$$

$$1-F(t) = e^{-\int_0^t \mu_x(s) ds} \quad \Rightarrow$$

$$F(t) = 1 - e^{-\int_0^t \mu_x(s) ds} \quad (X.1),$$

$$f(t) = (1-F(t)) \cdot \mu_x(t) = \mu_x(t) \cdot e^{-\int_0^t \mu_x(s) ds} \quad (X.2),$$

$$E(T) = \int_0^\infty t \cdot \mu_x(t) \cdot e^{-\int_0^t \mu_x(s) ds} \cdot dt = \int_0^\infty e^{-\int_0^t \mu_x(s) ds} dt.$$

Intensiteetin avulla voidaan siis lausua jäljellä olevan ajan jakauma (iässä x) ja tiheysfunktio. Mikäli $\mu_x(t)$ on vakio, niin tällöin saadaan $F(t) = 1 - e^{-\mu_x t}$. Tällöin T :n jakauma on eksponenttijakauma. Yleisesti intensiteetti ei kuitenkaan ole vakio, jolloin esitys (X.1) määrää T :n jakauman. Tarkastellaan vielä ehdollista todennäköisyyttä, kun $t > s$

$$1-F(t|T>s) = P(T>t|T>s) = \frac{P(T>t)}{P(T>s)} = \frac{1-F(t)}{1-F(s)} = \frac{e^{-\int_0^t \mu_x(s) ds}}{e^{-\int_0^s \mu_x(s) ds}} = e^{-\int_s^t \mu_x(s) ds}.$$

X.3 Ajan T kuluessa tapahtuvat suoritukset

Tarkastellaan seuraavaksi jäljellä olevana aikana eläkkeellä olevalle henkilölle maksettavia suorituksia (finanssimatematiikan merkinnät ks. Bowers ym. [1986]). Olkoon korkoutuvuus vakio δ . Ajan hetkellä $T = t$ suoritettava yksikön kokoinen kertosuoritus diskontattuna alkuun saadaan satunnaismuuttujan T muunnoksesta $k(T)$

$$k(T) = \begin{cases} 0 & , T < t \\ e^{-\delta \cdot T} & , T = t \\ 0 & , T > t. \end{cases}$$

Tällöin saadaan suoritettavan määrän odotusarvoksi

$$\begin{aligned}
E[k(T)] &= \int_0^{\infty} e^{-\delta \cdot t} \cdot f(t) dt \\
&= \int_0^{\infty} e^{-\delta \cdot t} \cdot (1 - F(t)) \cdot \frac{f(t)}{(1 - F(t))} dt \\
&= \int_0^{\infty} e^{-\delta \cdot t} \cdot (1 - F(t)) \cdot \mu_x(t) dt && \text{(X.3) ja Lemman X.1 mukaan} \\
&= 1 - \delta \cdot \int_0^{\infty} e^{-\delta \cdot t} \cdot (1 - F(t)) dt,
\end{aligned}$$

joista (X.3) on perinteinen vakuutusmatemaattinen esitystapa.

Yleensä esimerkiksi eläkettä ei suoriteta kerralla tietyn ajanjakson päättyessä vaan kaiken aikaa kunnes tila päättyy. Tarkkaan ottaen eläkettä maksetaan kerran kuussa yleensä kuun ensimmäisenä päivänä. Vakuutusmatematiikassa diskreetin suorituksen sijasta usein (mm. TEL:n laskuperusteissa) tarkastellaan ns. jatkuvaa suoritusta. Tämä tarkoittaa käytännössä sitä, että suoritetaan "jatkuvasti" määrä, jonka diskontattu arvo on kuten yllä

$$k(T) = e^{-\delta \cdot T}.$$

Tällöin suoritusta maksettavana aikana hetkeen T mennessä kaikenkaikkiaan suoritettavaksi määräksi saadaan

$$y(T) = \int_0^T k(s) ds = \int_0^T e^{-\delta \cdot s} ds = \frac{1 - e^{-\delta \cdot T}}{\delta} = \bar{a}_{\overline{T}|},$$

jolloin päättymishetkeen mennessä kaiken kaikkiaan suoritettavan määrän odotusarvoksi, pääoma-arvon $\bar{a}_{\overline{T}|}$ odotusarvoksi, saadaan

$$\begin{aligned}
E[y(T)] &= \int_0^{\infty} \bar{a}_{\overline{T}|} \cdot f(t) dt \\
&= \int_0^{\infty} \bar{a}_{\overline{T}|} \cdot (1 - F(t)) \cdot \mu_x(t) dt && \text{ja Lemman X.1 avulla} \\
&= \int_0^{\infty} e^{-\delta \cdot t} \cdot (1 - F(t)) dt && \text{(X.4)} \\
&= \int_0^{\infty} e^{-\delta \cdot t - \int_0^t \mu_x(s) ds} dt, && \text{joka seuraa (X.1):stä.}
\end{aligned}$$

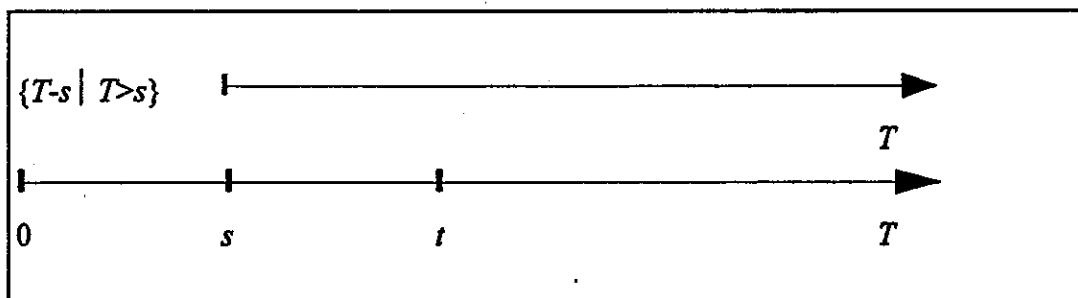
Selvästi saadaan myös $E[k(T)] + \delta \cdot E[y(T)] = 1$. Yllä olevien esitysten avulla saadaan siis laskettuksi odotusarvo $E[y(T)]$ tapahtumalle, että x -ikäiselle henkilölle maksetaan yksikön suuruista etuutta koko sen ajan, kun hän on tilassa $Z=a$. Esitys (X.4) on perinteinen

vakuutusmatemaattinen esitys vastaavasta odotusarvosta.

Tarkastellaan vielä tilannetta, jossa henkilö on pysynyt jo ajan s tilassa $Z=a$.

Tarkastelemme ehdollista odotusarvoa ehdolla $T > s$. Kuviossa X.1 on havainnollistettu mitä tämä tarkoittaa jäljellä olevan ajan kannalta.

Kuvio X.1



Nyt saamme jäljellä olevan ajan ehdolliseksi odotusarvoksi ehdolla $T > s$

$$\begin{aligned} E(T-s | T > s) &= \int_0^{\infty} (t-s) \cdot f(t | T > s) dt \\ &= \int_s^{\infty} (t-s) \cdot \frac{f(t)}{1-F(s)} dt \end{aligned}$$

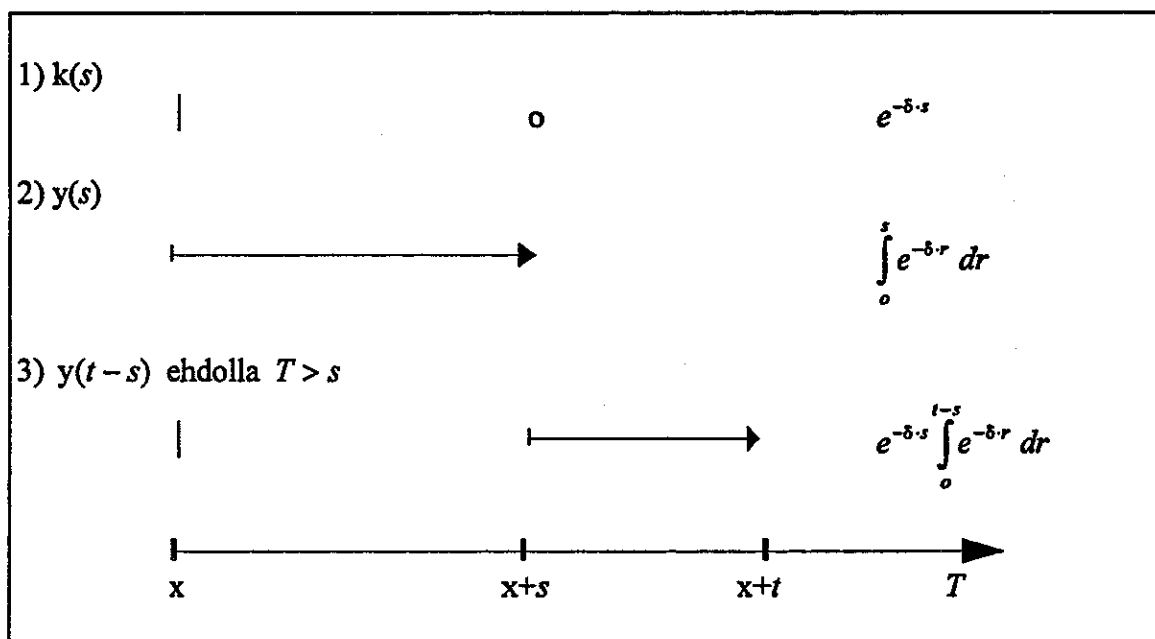
Vielä jäljellä olevana aikana kaiken kaikkiaan suoritettavaksi määräksi saadaan

$$\begin{aligned} E(y(T-s) | T > s) &= \int_0^{\infty} \bar{a}_{t-s} \cdot f(t | T > s) dt \\ &= \int_s^{\infty} \frac{1 - e^{-\delta(t-s)}}{\delta} \cdot f(t) dt \\ &= \frac{\int_s^{\infty} (1 - e^{-\delta(t-s)}) \cdot f(t) dt}{1 - F(s)} \\ &= \frac{\int_s^{\infty} e^{-\delta(t-s)} \cdot (1 - F(t)) \cdot dt}{(1 - F(s))}, \end{aligned} \tag{X.5}, \text{Lemmasta X.1}$$

$$\begin{aligned} &= \frac{\int_s^{\infty} e^{-\delta(t-s)} \cdot e^{-\int_0^t \mu_x(h) dh} dt}{e^{-\int_0^s \mu_x(h) dh}} \\ &= \int_s^{\infty} e^{-\delta(t-s)} \cdot e^{-\int_s^t \mu_x(h) dh} dt \end{aligned} \tag{X.6}$$

Kyseessä on odotusarvo kaiken kaikkiaan henkilölle maksettavalle edulle ikään $x+s$. Mahdolliset suoritukset iästä x ikään $x+s$ eivät nyt ole kiinnostavia, oleellista on vain se, että ollaan selviydytty ikään $x+s$ tilassa $Z=a$ ilman päättymistä. Kerätään vielä kuvioon X.2 kaikista kolmesta tapauksesta suoritettavat määrät diskontattuna ikään x .

Kuvio X.2



X.4 Satunnaismuuttujan T ja ajanjakson päättävän tapahtuman J yhteisjakauma

Siirrytään seuraavaksi tarkastelemaan tilannetta, jossa tilan päättävällä tapahtumalla on vaikutusta siihen minkälaisen edun piirissä henkilö on eli mitä hänelle maksetaan. Kiinnostuksen kohteena on tilanne, jossa henkilön on oltava tilassa $Z = a$, jotta hän siirtyessään vain syyn $J = i$ takia pois tilasta $Z = a$, olisi oikeutettu johonkin etuuteen, jota maksetaan jatkuvasti kunnes tämä uusi tila $Z = d$ päättyy.

Hajotetaan nyt tilan $Z=a$ päättymisen eri syihin $J = j, j = 1, \dots, k$, ja tarkastellaan yhdessä satunnaismuuttujia T ja J sekä niiden yhteisjakauman tiheysfunktiota $f(t, j)$.

Satunnaismuuttujien reunajakaumat ovat tällöin

$$f_T(t) = f(t) = \sum_{j=1}^k f(t, j) \text{ ja } f_J(j) = \int_0^\infty f(t, j) dt \text{ sekä tietysti } \int_0^\infty \sum_{j=1}^k f(t, j) dt = 1.$$

On huomattava, että päättymissytyt muodostavat osituksen mistä tietysti seuraa, että

$$\sum_{j=1}^k f_j(j) = 1 \text{ ja että } P_j(J = j, J = i) = 0.$$

On edelleen huomattava, että J on diskreetti satunnaismuuttuja ja tällöin tarkkaan ottaen pitäisi puhua sen pistetodennäköisyysfunktioista tai pistetodennäköisyysfunktion ja tiheysfunktion sekoituksesta. Nyt T :n reunakertymäfunktio

$$F_T(t) = F(t) = \int_0^t f_T(s) ds = \int_0^t \sum_{j=1}^k f(s, j) ds = \sum_{j=1}^k \int_0^t f(s, j) ds = \sum_{j=1}^k F(t, j),$$

eri päättymissyiden yhteiskertymäfunktioiden $F(t, j)$ summana saadaan T :n reunakertymäfunktio. Määritellään vielä syystä $J=j$ johtuvan päättymisen intensiteetti.

Määritelmä X.2

Olkoon $\mu_x^j(t) = \lim_{\Delta t \rightarrow 0+} \frac{P(t \leq T < t + \Delta t, J = j | T \geq t)}{\Delta t}$ päättävyyden intensiteetti syyhyn j ,
 jolloin $\mu_x^j(t) = \frac{f(t, j)}{1 - F(t)} = \frac{f(t, j)}{1 - \int_0^t \sum_{j=1}^k f(s, j) ds}$.

Tällöin kokonaispäättävyydeksi $\mu_x^S(t) = \mu_x(t)$ saadaan

$$\mu_x^S(t) = \mu_x(t) = \frac{f(t)}{1 - F(t)} = \frac{\sum_{j=1}^k f(t, j)}{1 - F(t)} = \sum_{j=1}^k \frac{f(t, j)}{1 - F(t)} = \sum_{j=1}^k \mu_x^j(t),$$

kokonaispäättävyys $\mu_x^S(t)$ on eri syiden päättävyyksien summa.

Koska nyt ajanjakson päättymisen tapahtuu useammasta eri tilasta täytyy ottaa käyttöön uusi merkintä. Merkinnällä $\mu_{x(a)}^j(t)$ tarkoitetaan intensiteettiä päättävyydelle, kun tila $Z = a$ päättyy ajanhetkellä $T = t$ syyhyn $J = j$.

Tarkastellaan seuraavaksi yhteistiheysfunktioita $f(t, j)$. Nyt

$$f(t, j) = \frac{f(t, j)}{(1 - F(t))} \cdot (1 - F(t)) = \mu_x^j(t) \cdot e^{-\int_0^t \mu_x(s) ds} = \mu_x^j(t) \cdot e^{-\int_0^t \sum_{j=1}^k \mu_x^j(s) ds}. \quad (X.7)$$

X.5 Edut, kun päättymissyyt vaikuttavat

Yhteistiheysfunktiolla pystytään laskemaan etuja, jos tila $Z=a$ päättyy tiettyyn syyhyn $J=j$. Tarkastellaan etua, jolla on arvo vain, jos tilan päättävä syy on tietty esimerkiksi $J=i$. Mikäli ajanjakso päättyy muuhun syyhyn, edulla ei ole arvoa. Tällöin yksikön suuruisen kertosuorituksen arvo saadaan odotusarvona satunnaismuuttujien T ja J muunnoksesta $k(T, J)$, kun ajatellaan, että ajanjakso päättyy ajassa t

$$k(T, J) = \begin{cases} J=i & J \neq i \\ 0 & 0 & , T < t \\ e^{-\delta \cdot T} & 0 & , T = t \\ 0 & 0 & , T > t. \end{cases}$$

Nyt voimme laskea kuten aiemmin edun odotusarvon.

$$\begin{aligned} E[k(T, J)] &= \int_0^{\infty} \sum_{j=1}^k k(t, j) \cdot f(t, j) dt \\ &= \int_0^{\infty} e^{-\delta \cdot t} \cdot f(t, i) dt. \end{aligned}$$

Ajatellaan nyt lisäksi, että tilan $Z = a$ päättymisen syyllä $J=i$, johtaa henkilön toiseen tilaan $Z = d$ hetkellä s , jossa aletaan esimerkiksi maksaa jatkuvaa suoritusta kunnes tämä uusi tila $Z = d$ päättyy. Tilassa $Z = d$ tapahtuvaa suoritusta vastaava ehdollinen odotusarvo $E(y(T-s) | T > s)$ on jo aiemmin laskettu, suorituksen alkamisen ehtona on paitsi, että ollaan tilassa $Z = d$ myös, että ollaan ohitettu jo ajanjakso $T > s$. Kaikenkaikkiaan maksettava määrä on

$$z(T, J) = \begin{cases} J=i & J \neq i \\ 0 & 0 & , T < s \\ e^{-\delta \cdot s} \cdot \frac{1}{\delta} & 0 & , T \geq s. \end{cases}$$

Tällöin saamme ehdollisen odotusarvon ominaisuuksista

$$\begin{aligned} E[z(T, J)] &= \int_0^{\infty} f(s) \cdot E(z(T, J) | T > s) ds \\ &= \int_0^{\infty} \sum_{j=1}^k f(s, j) \cdot E(z(T, J) | T > s) ds = \int_0^{\infty} \sum_{j=1}^k f(s, j) \cdot \int_0^{\infty} \sum_{j=1}^k f(t, j | T > s) z(t, j) ds dt \end{aligned}$$

$$\begin{aligned}
&= \int_0^\infty f(s, i) \cdot \int_0^\infty f(t, i | T > s) \cdot e^{-\delta \cdot s} \cdot \bar{a}_{t-s} ds dt = \int_0^\infty f(s, i) \cdot e^{-\delta \cdot s} \cdot E(y(T-s) | T > s) ds \\
&= \int_0^\infty (1 - F(s)) \cdot \frac{f(s, i)}{(1 - F(s))} \cdot e^{-\delta \cdot s} \cdot \left[\int_{s+}^\infty e^{-\delta \cdot (t-s)} \cdot (1 - F(t | T > s)) dt \right] ds \quad (X.8)
\end{aligned}$$

$$\begin{aligned}
&= \int_0^\infty e^{-\int_0^s \mu_{x(a)}(h) dh} \cdot \mu_{x(a)}^i(s) \cdot e^{-\delta \cdot s} \cdot \left[\int_{s+}^\infty e^{-\delta \cdot (t-s)} \cdot e^{-\int_s^t \mu_{x(a)}(h) dh} dt \right] ds \\
&= \int_0^\infty e^{-\delta \cdot s} \cdot e^{-\int_0^s \mu_{x(a)}^i(h) dh} \cdot \mu_{x(a)}^i(s) \cdot \left[\int_{s+}^\infty e^{-\delta \cdot (t-s)} \cdot e^{-\int_s^t \mu_{x(a)}(h) dh} dt \right] ds \quad (X.9).
\end{aligned}$$

X.6 Mallin varianssit

Varianssit voidaan laskea kaavan $VAR(X) = E(X^2) - E(X)^2$ avulla. Odotusarvot on jo yllä laskettu, pitää tarkastella vain satunnaismuuttuja X :n muunnosta $Y = X^2$. Kun sovelletaan Lemmaa X.1, niin tarvitaan

$$D_t \left(\frac{1 - e^{-\delta \cdot (t-s)}}{\delta} \right)^2 = 2 \cdot \left(\frac{1 - e^{-\delta \cdot (t-s)}}{\delta} \right) \cdot e^{-\delta \cdot (t-s)} = 2 \cdot e^{-\delta \cdot (t-s)} \cdot \bar{a}_{t-s} = 0, \text{ kun } t = s.$$

joten $y^2(T-s)$ odotusarvo ehdolla $T > s$ voidaan lausua muodossa

$$\begin{aligned}
E(y^2(T-s) | T > s) &= \int_0^\infty \left(\frac{1 - e^{-\delta \cdot t}}{\delta} \right)^2 \cdot f(t | T > s) dt \\
&= \int_{s+}^\infty 2 \cdot e^{-\delta \cdot (t-s)} \cdot \bar{a}_{t-s} \cdot e^{-\int_s^t \mu_{x(a)}(h) dh} dt \text{ ja toisaalta}
\end{aligned}$$

$$\begin{aligned}
E[z^2(T, J)] &= \int_0^\infty f(s) \cdot E(z^2(T, J) | T > s) ds = \int_0^\infty f(s, i) \cdot e^{-2\delta \cdot s} \cdot E(y^2(T-s) | T > s) ds \\
&= \int_0^\infty e^{-2\delta \cdot s} \cdot e^{-\int_0^s \mu_{x(a)}^i(h) dh} \cdot \mu_{x(a)}^i(s) \cdot \left[\int_{s+}^\infty 2 \cdot e^{-\delta \cdot (t-s)} \cdot \bar{a}_{t-s} \cdot e^{-\int_s^t \mu_{x(a)}(h) dh} dt \right] ds.
\end{aligned}$$

Variansseja tarvitaan, kun esimerkiksi lasketaan etujen odotusarvoille luottamusvälejä.

Näiden sekä otantateorian ja keskeisen raja-arvolauseen avulla voidaan konstruoida pääoma-arvoille laskuperusteiden toimivuuden kannalta tietyissä tilanteissa välttämättömiä varmuusvaroja.

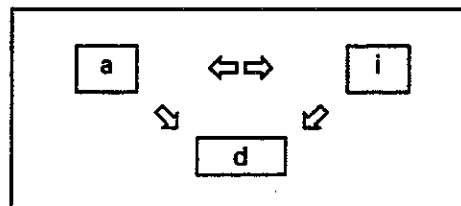
Y. Tilaluokkamallit

Y.1. Taustaa

Viime vuosina työkyvyttömyyden pääoma-arvoja on käsitelty ns. tilaluokkamallien avulla. Tilayhtälömallia on kehitelty useassa eri lähteessä. Lähestymme tässä mallin johtamista lähinnä Pitaccon [1995] ja Wolthuisin [1994] mukaan. Luvussa X tarkasteltiin jäljellä olevaa aikaa T , jonka päättää syy $J=j$, kun ollaan tilassa $Z=z$. Nyt tarkastelemme satunnaismuuttujaa Z tarkemmin ja luovumme oletuksesta, että aika T kuluu tietyssä tilassa $Z=z$ kaiken aikaa.

Tarkastelemme stokastista prosessia $\{Z(t), t \geq 0\}$; jatkuvaa epähomogeenistä semi-Markov prosessia, jossa on kolme tilaa: $Z=a$ (aktiivi), $Z=i$ (työkyvytön) ja $Z=d$ (kuollut). Tila $Z=d$ on absorboiva. Kuviossa Y.1. on kuvattu tilojen väliset yhteydet.

Kuvio Y.1



Kuten kuvioista huomataan tilayhtälömallissa hyväksytään siirtymiset tilojen $Z=a$ ja $Z=i$ välillä. Yleensä stokastisten prosessien ja vakuutusmatematiikan yhteydessä tilojen välisten siirtymisten todennäköisyyksiä eli ns. transitiotodennäköisyyksiä merkitään seuraavasti $p_{ij}(s, t) = P(Z(x+t) = j | Z(x+s) = i) = P(T > t; \exists u \in (s, t] \text{ s.e. } Z = j \mid \forall v \in [u, t] | Z = i; T > s)$. Kuten huomaamme tilaan $Z=j$ voidaan joutua minä hetkenä hyvänsä välillä $[0, t]$ eli emme aivan tarkaan tiedä (emmekä välitä tietää) koska tilaan $Z=j$ siirrytään, joten otetaan huomioon kaikki mahdolliset hetket. Emme myöskään tiedä kuinka monta kertaa mennään mallin ei-absorboivien tilojen välillä edestakaisin, joten otamme huomioon kaikki mahdolliset siirtymäkertamäärät. Kuitenkin tiedämme varmasti, että lopulta hetkellä $T=t$ olemme tilassa $Z=j$.

Merkitään edelleen $p_{ii}(t, 0) = P(T > t; \forall u \in (0, t] \text{ s.e. } Z = i | Z = i; T > 0)$. Tätä todennäköisyyttä laskiessa emme salli tilanvaihtoja tilojen $Z=i$ ja $Z=a$ välillä.

Todennäköisyys on identtinen luvun X.2 elossaolotodennäköisyyden $S(t) = 1 - F(t)$ kanssa. Luvun X mukaan jäljellä oleva aika kuluu kokonaan tietyssä tilassa $Z=i$ kuten tässä.

Y.2. Tilaluokkamallin johtaminen

Määrittelemme aluksi ns. *transitiointensiteetit*.

Määritelmä Y.1

Olkoon $\mu_x^y(s) = \lim_{t \rightarrow s} \frac{P(Z(x+t)=j | Z(x+s)=i)}{t-s} = \lim_{t \rightarrow s} \frac{p_{ij}(s,t)}{t-s}$ *transitiointensiteetti* ajan hetkellä s .

Markov-ominaisuutta käyttäen voidaan johtaa ns. Chapman-Kolmogorov -yhtälö.

$$\begin{aligned} p_{ij}(s,t) &= P(Z(x+t)=j | Z(x+s)=i) \\ &= \sum_{k \in N} P(Z(x+t)=j; Z(x+u)=k | Z(x+s)=i) \\ &= \sum_{k \in N} P(Z(x+u)=k | Z(x+s)=i) \cdot P(Z(x+t)=j | Z(x+u)=k) \\ &= \sum_{k \in N} p_{ik}(s,u) \cdot p_{kj}(u,t), \quad 0 \leq s \leq u \leq t, N = \{a, i, d\}. \end{aligned}$$

Summaan riittää laskea yhteen mallin ei-absorboivat tilat. Laskettaessa pysymistä tietyssä tilassa tai siirtymistä tilojen välillä on $P(Z(x+t)=j | Z(x+s)=d) = 0, \forall t > s$. Chapman-Kolmogorov yhtälön avulla saadaan Kolmogorovin forward-differentiaaliyhtälöt

$$\begin{aligned} \frac{d}{dt} p_{aa}(s,t) &= -p_{aa}(s,t) \cdot (\mu_x^{ai}(t) + \mu_x^{ad}(t)) + p_{ai}(s,t) \cdot \mu_x^{ia}(t) \\ \frac{d}{dt} p_{ai}(s,t) &= -p_{ai}(s,t) \cdot (\mu_x^{ia}(t) + \mu_x^{id}(t)) + p_{aa}(s,t) \cdot \mu_x^{ai}(t) \\ \frac{d}{dt} p_{ad}(s,t) &= p_{aa}(s,t) \cdot \mu_x^{ad}(t) + p_{ai}(s,t) \cdot \mu_x^{id}(t) \\ \frac{d}{dt} p_{ia}(s,t) &= -p_{ia}(s,t) \cdot (\mu_x^{ai}(t) + \mu_x^{ad}(t)) + p_{ii}(s,t) \cdot \mu_x^{ia}(t) \\ \frac{d}{dt} p_{ii}(s,t) &= -p_{ii}(s,t) \cdot (\mu_x^{ia}(t) + \mu_x^{id}(t)) + p_{ia}(s,t) \cdot \mu_x^{ai}(t) \\ \frac{d}{dt} p_{id}(s,t) &= p_{ii}(s,t) \cdot \mu_x^{id}(t) + p_{ia}(s,t) \cdot \mu_x^{ad}(t). \end{aligned} \tag{Y.1}$$

Yhtälöjoukon (Y.1) ratkaiseminen vaatii numeerisia menetelmiä. Edelleen saamme

$$\begin{aligned}\frac{d}{dt} p_{aa}(s, t) &= -p_{aa}(s, t) \cdot (\mu_x^{ai}(t) + \mu_x^{ad}(t)) \\ \frac{d}{dt} p_{ii}(s, t) &= -p_{ii}(s, t) \cdot (\mu_x^{ia}(t) + \mu_x^{id}(t)).\end{aligned}\quad (Y.2)$$

Pitacco [1995] johtaa todennäköisyydelle $p_{ai}(0, t)$ vielä uuden esitysmuodon:

$$p_{ai}(0, t) = \int_0^t p_{aa}(0, u) \cdot \mu_x^{ai}(u) \cdot p_{ii}(u, t) du. \quad (Y.3)$$

ja tämän avulla hän saa (Lemma X.1 käyttäen)

$$\bar{a}_x^{ai} = \int_0^\infty e^{-\delta \cdot t} \cdot p_{ai}(0, t) dt = \int_0^\infty e^{-\delta \cdot t} \cdot \int_0^t p_{aa}(0, u) \cdot \mu_x^{ai}(u) \cdot p_{ii}(u, t) dt du, \quad (Y.4)$$

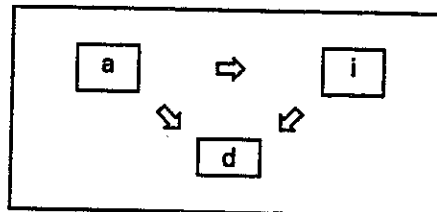
josta vaihtamalla integrointijärjestys saadaan

$$\bar{a}_x^{ai} = \int_0^\infty e^{-\delta \cdot u} \cdot p_{aa}(0, u) \cdot \mu_x^{ai}(u) \cdot \left[\int_u^\infty e^{-\delta \cdot (t-u)} \cdot p_{ii}(u, t) dt \right] du. \quad (Y.5)$$

Merkintä $\bar{a}_x^{ai}$ tarkoittaa vastaisen työkyvyttömyyseläkkeen pääoma-arvoa. Esitys (Y.5) on jo tuttu luvusta X.4. Itseasiassa se on melkein identtinen esitykselle (X.8). Erona on todennäköisyyden $p_{aa}(0, u)$ sekä $\mu_x^{ai}(t)$ käsittely. Luvussa X.4 ei sallittu ennen työkyvyttömyyden alkamista edestakaista siirtymistä tilojen $Z=a$ ja $Z=i$ välillä. Tarkastelemme tätä seikkaa hieman tarkemmin.

Jos emme halua ottaa huomioon tilojen $Z=a$ ja $Z=i$ välisiä tilansiirtoja ja oletamme, että alussa olemme tilassa $Z=a$, emme voi sallia siirtymistä tilasta $Z=i$ takaisin tilaan $Z=a$. Kuvio Y.2 selventää tilayhtälömallia.

Kuvio Y.2



Tällöin pääoma-arvoa laskettaessa (Y.4) todennäköisyyden $p_{aa}(0, u)$ sijasta on yksinkertaisesti käytettävä yhtälön (Y.2) avulla saatavaa $p_{aa}(0, u)$, jolloin (Y.5) ja (X.8) vastaavat toisiaan, kunhan myös transitiointensiteetti $\mu_x^{ij}(t)$ vastaa luvun X mukaista

päättymiseen $J=j$ liittyvää intensiteettiä $\mu_x^j(t)$. Tämä on uskottavaa, koska (Y.3) pystytään kirjoittamaan muotoon

$$p_{ai}(0,t) = \int_0^t p_{aa}(0,u) \cdot \mu_x^{ai}(u) \cdot p_{ii}(u,t) du \quad (Y.6)$$

$$= \int_0^t (1 - F(u, Z=i)) \cdot \frac{f(u, Z=j)}{1 - F(u, Z=i)} \cdot \frac{1 - F(t, Z=j)}{1 - F(u, Z=j)} du, \quad (Y.7)$$

joka vastaa luvussa X esitettyä mallia (X.8). Esityksen (Y.7) mukainen tilayhtälömalli on otettu Suomessa käytettävän ns. Z-mallin pohjaksi.

Y.3. Kaksi approksimaatiota

Tarkastellaan seuraavaksi kahta approksimaatiota todennäköisyydelle $p_{aa}(0,u)$.

Approksimaatio 1

Differentiaaliyhtälön (Y.1) mukaan

$$\frac{d}{dt} p_{aa}(s,t) = -p_{aa}(s,t) \cdot (\mu_x^{ai}(t) + \mu_x^{ad}(t)) + p_{ai}(s,t) \cdot \mu_x^{ia}(t).$$

Yksi intuitiivinen approksimaatio saadaan, jos kirjoitetaan

$$\frac{d}{dt} p_{aa}(s,t) \approx -p_{aa}(s,t) \cdot (\mu_x^{ai}(t) + \mu_x^{ad}(t) - \mu_x^{ia}(t)). \quad (Y.8)$$

Jotta näin voidaan tehdä, täytyy siirtymätodennäköisyys arvioida: $p_{ai}(s,t) \approx p_{aa}(s,t)$. Tämä tarkoittaa suunnilleen sitä, että henkilön - aktiivina ollessaan - intensiteetti kuolla ja sairastua on samansuuruinen kuin - työkyvyttömänä ollessaan - intensiteetti kuolla ja parantua:

$$\begin{aligned} p_{aa}(s,t) &= p_{ai}(s,t) \Rightarrow \\ \frac{d}{dt} p_{aa}(s,t) &= \frac{d}{dt} p_{ai}(s,t) \Leftrightarrow \\ -p_{aa}(s,t) \cdot (\mu_x^{ai}(t) + \mu_x^{ad}(t)) + p_{ai}(s,t) \cdot \mu_x^{ia}(t) \\ &= -p_{ai}(s,t) \cdot (\mu_x^{ia}(t) + \mu_x^{id}(t)) + p_{aa}(s,t) \cdot \mu_x^{ai}(t) \\ 2 \cdot \mu_x^{ai}(t) + \mu_x^{ad}(t) &= 2 \cdot \mu_x^{ia}(t) + \mu_x^{id}(t), \end{aligned}$$

jolloin voidaan kirjoittaa

$$\frac{d}{dt} p_{aa}(s, t) = -p_{aa}(s, t) \cdot (\mu_x^{ai}(t) + \mu_x^{ad}(t) - \mu_x^{ia}(t)), \text{ mutta nyt on huomattava, että}$$

$$p_{aa}(s, t) = 1 - e^{-\int_0^t (\mu_x^{ai}(u) + \mu_x^{ad}(u) - \mu_x^{ia}(u)) du} \quad (\text{Y.9}).$$

Tässä mallissa parantuminen otetaan suoraan sairastumisen ja kuoleamisen vähennykseksi. Mallin sovittaminen on myös suoraviivaista: otetaan huomioon tarkasteluaikana parantuneet henkilöt.

Approksimaatio 2

Toinen mahdollisuus aproksimoida todennäköisyyttä $p_{aa}(s, t)$ on sallia siirtymiset tilojen $Z=a$ ja $Z=i$ välillä, mutta vain rajoitetusti, esimerkiksi yhden kerran ennen lopullista sairastumista. Ositetaan tapahtuma siirtymiskertojen määrän avulla. Tällöin

$$p_{aa}(0, t) = p_{aa}(0, t) + \int_0^t p_{aa}(0, u) \cdot \mu_x^{ai}(u) \cdot \left[\int_u^t p_{ii}(u, v) \cdot \mu_x^{ia}(v) \cdot p_{aa}(v, t) dv \right] du + ..$$

Tällöin ensimmäisen siirtymisen intensiteetti on oletuksen mukaan erilainen kuin toisen siirtymisen intensiteetti eli tiheysfunktioita vastaavaa termiä kirjoittaessa on merkittävä

$$\frac{d}{dt} p_{aa}(0, t) = \mu_x^{ai}(u) \cdot p_{aa}(0, t) + \mu_x^{ai^2}(u) \cdot \int_0^t p_{aa}(0, u) \cdot \mu_x^{ai}(u) \cdot \left[\int_u^t p_{ii}(u, v) \cdot \mu_x^{ia}(v) \cdot p_{aa}(v, t) dv \right] dt$$

$$+ \dots$$

missä $\mu_x^{ai^2}(u)$ tarkoittaa intensiteettiä sairastua toisen kerran ja tässä nyt $\mu_x^{ai}(u)$ vastaavasti intensiteettiä sairastua ensimmäisen kerran. Vastaavasti mallia voidaan tarkentaa laskemalla todennäköisyydet $p_{aa}(s, t)$ niin, että ollaan tilassa joko ensimmäistä tai toista kertaa - tai tietysti voidaan arvioida ettei tällä ole relevanssia.

On pääteltävä havainnoista kuinka monta kertaa mallissa on sallittava tilansiirrot. Myös nyt intensiteetit saadaan sovitettua malliin kuten luvussa X, kunhan erotellaan ne tapaukset, joissa sairastuminen tapahtuu ensimmäisen kerran ja toisen (tai useamman) kerran.

Z. Z-malli

Z.1 Taustaa

Suomessa työntekijäin eläkelain mukaisessa eläkevakuutuksessa (TEL) työkyvyttömyyden pääoma-arvot ja tarvittavat todennäköisyydet on laskettu käyttäen hyväksi ns. Z-mallia. Z-mallia ovat käsitelleet SHV-tutkielmissaan mm. Lassila [1977], Gauffin [1981] ja Tuomikoski [1986] sekä Gauffin ja Leinonen [1982].

Z.2 Z-mallin todennäköisyydet

TEL:n mukaisen vakuutuksen yleisissä laskuperusteissa on määritelty kuinka työkyvyttömyyden pääoma-arvot on laskettava. Niiden mukaan integraali

$$\int_{u_1}^{u_2} z(t, u) du$$

ilmoittaa todennäköisyyden sille, että vastasyntynyt on elossa ajan t kuluttua ja on tällöin ollut yhdenjaksoisesti työkyvytön ajanjakson, jonka pituus on välillä (u_1, u_2) . Tällöin $z(t, u)$ -funktiossa oletetaan, että vastasyntynyt on elossa iässä t , hän tulee työkyvyttömäksi iässä $t-u$ ja on tämän jälkeen yhtäjaksoisesti työkyvytön ikään t eli ajan u .

Huomautettakoon, että sanallisesti z -funktiossa ennen ajassa $t-u$ tapahtuvaa sairastumista edeltävään aikaan ei liity muuta oletusta kuin, että vastasyntynyt on elossa. Teoriassa hän voisi siis käydä välillä työkyvyttömänä ja tulla taas terveeksi kunhan pysyy elossa. Oletus siitä, että elossa oleminen pitää sisällään vain terveenä olemista, on seurausta valitusta elossaolomallista ennen työkyvyttömyyden alkamista. On huomattava, että z -funktio on määritelty vain vastasyntyneelle, joten kuhunkin ikäluokkaan $x > 0$ liittyvät todennäköisyydet on saatava yhden funktion avulla.

Nyt z -funktio vastaa (Y.6):n integraalin sisällä olevaa esitystä, kun huomataan että $x=0$

$$z(t, u) = p_{aa}(0, t-u) \cdot \mu_0^{ai}(t-u) \cdot p_{ii}(t-u, t).$$

Mallin (X.5) mukainen todennäköisyys saadaan kunhan huomataan, että

$$\begin{aligned} \frac{z(t+h, u+h)}{z(t, u)} &= \frac{p_{aa}(0, t-u) \cdot \mu_0^{ai}(t-u) \cdot p_{ii}(t-u, t+h)}{p_{aa}(0, t-u) \cdot \mu_0^{ai}(t-u) \cdot p_{ii}(t-u, t)} \\ &= \frac{p_{ii}(t-u, t+h)}{p_{ii}(t-u, t)} = p_{ii}(t, t+h) \end{aligned} \quad (Z.1)$$

Koska z -funktio kuvaa sanallisesti koko yllä kuvattua tapahtumaketjua, ei siitä ole suoraan irrotettavissa erikseen elossa olemiseen liittyvää todennäköisyyttä. Jotta pystyisimme tarkastelemaan iässä $x > 0$ mallin (Y.4) mukaista pääoma-arvoa (jotka liittyvät vastaisiin eläketapahtumiin) lasketaan

$$\begin{aligned} \frac{z(t,u)}{p_{aa}(0,x)} &= \frac{p_{aa}(0,t-u) \cdot \mu_0^{ai}(t-u) \cdot p_{ii}(t-u,t)}{p_{aa}(0,x)} \\ &= p_{aa}(x,t-u) \cdot \mu_0^{ai}(t-u) \cdot p_{ii}(t-u,t) \end{aligned} \quad (Z.2),$$

joka vastaa sitä, että tarkasteluikä on nyt x ja vastasyntynyt on jo ohittanut aikavälin $[0,x)$.

Z.3 Z-mallin muoto ja sisältö

TEL:n mukaisen vakuutuksen yleisten laskuperusteiden mukaan arvoilla $t \geq u \geq 0$

$$\int_0^t z(t,u) du = e^{-a_4 \cdot t} \quad (Z.3)$$

ja arvoilla $t \geq u \geq e$

$$z(t,u) = \sum_{j=0}^2 z_j(t,u) = \sum_{j=0}^2 a_j \cdot e^{b_j \cdot t - c_j \cdot u} = \sum_{j=0}^2 e^{b_j \cdot (t-u)} \cdot a_j \cdot e^{-(c_j - b_j) \cdot u} \quad (Z.4)$$

Käytännössä e on lyhyin työkyvyttömyyden kesto. Työkyvyttömäksi tulemisen jälkeen työkyvyttömyyseläke ei ala heti, vaan sairausvakuutuksen kautta maksetaan päivärahaa yleensä yhdeksän kuukautta ($e=9/12$) ennen työkyvyttömyyseläkeen alkamista.

Tuomikoski [1986] huomauttaa ja tämä nähdään myös esityksen (Z.4) avulla

$$z_j(t+h, u+h) = e^{b_j \cdot (t+h-u-h)} \cdot a_j \cdot e^{-(c_j-b_j) \cdot (u+h)} = z_j(t,u) \cdot e^{-(c_j-b_j) \cdot h}. \text{ Nyt}$$

$$\frac{z(t+h, t+h)}{z(t,u)} = \sum_{j=0}^2 \frac{z_j(t,u)}{z(t,u)} \cdot e^{-(c_j-b_j) \cdot h}, \quad (Z.5)$$

jolloin jo alkaneen työkyvyttömyyseläkkeen *elossaolotodennäköisyytenä* käytetään kolmen eksponenttijakauman sekoitusta. Elosaolotodennäköisyyksien sekoitus riippuu myös z -mallista ja sen parametreista: nykyiästä t ja menneestä sairausajan kestosta u .

Vastaisen eli tulevaisuudessa alkavan eläkkeen pääoma-arvoa laskettaessa todennäköisyys säilyä hengissä syntymästä nykyikään x arvioidaan seuraavasti $p_{aa}(0,x) = e^{-a_4 \cdot x}$, joten

tarvittava mallin (X.8) ja (Y.4) mukainen todennäköisyys (Z.2) on, kun vielä merkitään $t = x + d$; on huomattava, että $d \geq u$

$$\frac{z(t,u)}{e^{-a_4 \cdot x}} = \sum_{j=0}^2 e^{b_j \cdot (t-u) + a_4 \cdot x} \cdot a_j \cdot e^{-(c_j - b_j) \cdot u} = \sum_{j=0}^2 e^{b_j \cdot (d-u) + (a_4 + b_j) \cdot x} \cdot a_j \cdot e^{-(c_j - b_j) \cdot u} \quad (Z.6)$$

Vastaisten eläkkeiden pääoma-arvossa jakaumien sekoitussuhde on määrittelemätön.

Z.4 Nykyisen Z-mallin sisällön tarkastelua

Alkaneen eläkkeen elossaolotodennäköisyytenä käytetään kolmen iästä riippumattoman jakauman sekoitusta. Sekoitus riippuu iästä ja sairauden menneestä kestoista, mutta varsinaiset jakaumat ovat iästä riippumattomia eksponenttijakaumia. Näiden jakaumien odotusarvot EU_j on helposti laskettavissa ja tulkittavissa tulevan sairausajan keston odotusarvoksi. Aivan samoin vakiokuolevuuteen a_4 liittyvä elossaolotodennäköisyyttä vastaava jakauma on eksponenttijakauma, joten myös tähän malliin liittyvä eliniän odotusarvo EX on laskettavissa. Vakiokuolevuuteen a_4 liittyväksi elinajan odotusarvoksi saadaan $EX = 1 / a_4 = 217$. Laskuperusteissa mallin käyttö kuitenkin rajoitetaan ikäalueeseen 0 - 70, joten ikäalueen rajoittaminen huomioiden mallin elinajan odotusarvo $EX = 61$, kun $x \leq 70$, kun $S(t) = 0$, kun $x > 70$. Taulukkoon 1 on koottu näiden jakaumien odotusarvot nykyisen perusteen mukaisena. Tarkasteluyksikkö on vuosi. Huomautettakoon, että työkyvyttömyyseläkkeelle voi joutua TEL:n mukaan vain ikävälillä 14 - 65 vuotta.

Taulukko 1

j	$EU_j = 1 / (c_j - b_j)$	$EX = 1 / a_4$	$EX, x \leq 70$
1	1,6	217	61
2	62,5		
3	20,0		

Vastaisten eläkkeiden puolella jakaumien tulkinta ei ole yhtä suoraviivainen. Gauffin ja Leinonen [1982] huomauttavat myös, että vastaisten eläkkeiden pääoma-arvoa laskettaessa Z-mallin (Z.4) työkyvyttömyyden elossaolotodennäköisyyden on oletettu noudattavan kussakin komponentissa j eksponenttijakaumaa $\exp(c_j - b_j)$. Tarkastellaan vastaista pääoma-arvoa hieman tarkemmin. Mikäli kaikki b_j kertoimet olisivat negatiivisia, olisi kuolevuus ennen työkyvyttömyyttä samoin kolmen eksponenttijakautuman sekoite ja työkyvyttömäksi jouduttaisiin samoilla iästä riippumattomilla siirtymäintensiteeteillä a_j . Työkyvyttömäksi jouduttaisiin tällöin kolmea erillistä reittiä niin, että kullakin ryhmällä

olisi omanlaisensa ketju, jonka muodostavat aktiivin elossaolotodennäköisyys, sairastumisintensiteetti ja sairastuneen elossaolotodennäköisyys. Tämä lähestymistapa olisi mielenkiintoinen, mikäli esimerkiksi sairastumisyyt olisivat jaoteltavissa kolmeen luokkaan niin, että kullakin luokalla olisi oma työkyvyttömyyden keston keskimääräinen pituus.

Koska nykyisen perusteen mukaiset b_j kertoimet ovat suurempia kuin nolla, on parempi tulkita nyt olemassaolevassa z -mallissa, että kertoimet b_j pitävät sisällään paitsi kuolevuutta ennen työkyvyttömyyttä myös intensiteettiä tulla työkyvyttömäksi. Tällöin voimme olettaa, että aktiivin elossaolotodennäköisyys ennen työkyvyttömyyttä - kuten laskuperusteissa on jo oletettu - saadaan eksponenttijakauman avulla vakiointensiteetillä a_4 . Tällöin $b_j = -a_4 + b_j^* \Rightarrow b_j^* = b_j + a_4$ ja (Z.6) voidaan kirjoittaa muodossa, kun muistetaan, että $t = x + d$

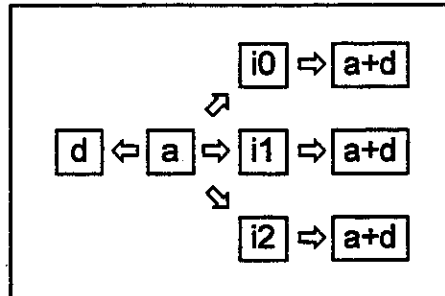
$$\sum_{j=0}^2 e^{(-a_4 + b_j^*) \cdot (t-u) + a_4 \cdot x} \cdot a_j \cdot e^{-(c_j - b_j) \cdot u} = \sum_{j=0}^2 e^{-a_4 \cdot (d-u)} \cdot \left[a_j \cdot e^{(b_j^*) \cdot (t-u)} \right] \cdot e^{-(c_j - b_j) \cdot u}$$

Tässä mallissa aktiivin elossaolotodennäköisyydellä ennen työkyvyttömyyttä on eksponenttijakaumatulkinta. Vastaavasti intensiteetillä joutua työkyvyttömäksi ei ole enää samaa tulkintaa, sillä se riippuu nyt kustakin sairastumisistä $t-u$ ja se onkin Gompertz-tyyppinen. Oletus, että aktiivin elossaolotodennäköisyys ennen työkyvyttömyyttä on eksponenttijakautunut, on itseasiassa pakkokin tehdä. Kun vastaisen eläkkeen pääoma-arvossa funktio $z(t,u)$ jaetaan termillä $e^{-a_4 \cdot x}$, diskontataan juuri tällä määrällä jo mennyttä aktiivin elossaolotodennäköisyyttä pois Z -mallin vastasyntyneeltä ja jäljelle jäävä määrä b_j -kertoimista on jotain muuta. On eri asia onko a_4 -intensiteetin mukainen eksponenttimalli realistinen ja toisaalta pitäisikö sitä soveltaa kuolevuuteen iästä x ikään $t = x + d$. Tietysti sitä sovelletaan ikäryhmän $y = x + d$ pääoma-arvoa laskettaessa. Samoin oletusta voidaan perustella Z -mallin tämän elossaolotodennäköisyyden estimointitavalla (Tuomikoski [1986] s.22). Näillä oletuksilla voidaan ylläoleva esitys kirjoittaa muodossa

$$\sum_{j=0}^2 e^{-a_4 \cdot (d-u)} \cdot \left[a_j \cdot e^{(b_j^*) \cdot (t-u)} \right] \cdot e^{-(c_j - b_j) \cdot u} = e^{-a_4 \cdot (d-u)} \cdot \sum_{j=0}^2 \left[a_j \cdot e^{(b_j^*) \cdot (t-u)} \right] \cdot e^{-(c_j - b_j) \cdot u}$$

mikä johtaa malliin, jota voidaan kuvata kuvion Z.1 avulla.

Kuvio Z.1



Kuviossa Z.1 on eroteltu eri tilat j olla työkyvyttömänä, jotta korostuisi, että sairastumisintensiteettejä on kolmenlaisia, kuten myös työkyvyttömyyden päättymisintensiteettejä. Tila $a+d$ tarkoittaa, että sairaanaolemisen voi päättää sekä kuolema että paraneminen, eivätkä ne ole intensiteetissä eroteltavissa.

Kuten luvusta X muistetaan, kokonaisintensiteetti oli osiensa summa. Kokonaissairastuvuudeksi iässä x saadaan

$$\mu_x(x) = \sum_{j=0}^2 \mu_x^j(x) = \sum_{j=0}^2 a_j \cdot e^{(b_j^*) \cdot x}, \quad \text{kun } (x = \lim_{u \rightarrow 0} (t - u)) \quad (\text{Z.7}).$$

Gauffin [1981], Gauffin ja Leinonen [1982] sekä Tuomikoski [1986] ovat antaneet oleellisesti saman sairastumisintensiteetin sen lisäksi, että he ovat ottaneet huomioon tarkastelussa osotusajan e ja ovat tutkineet eläkeyhtiön kannalta olennaisia työkyvyttömyyseläkkeen alkamisen intensiteettejä. Liitteen Z kuviossa Z3.1 on piirretty esityksen (Z.7) ja aiemman tutkimuksen (T.13) mukaiset sairastumisintensiteetit. Kuviot seuraavat toisiaan lukuunottamatta aivan ikäalueen loppupäätä.

On myös helposti todettavissa, että nykyperusteen mukaan vastaisen eläkkeen Z-mallin työkyvyttömyyden elossaolotodennäköisyyksien jakaumat eivät ole ilman painoja suoraan seurausta saman todennäköisyyskentän osituksesta. Tämä nähdään, koska

$$\exists u > 0 \text{ s.e. } \sum_{j=0}^2 e^{-(c_j - b_j) \cdot u} > 1. \quad (\text{Z.8})$$

Mikäli halutaan, että yo. summa on ≤ 1 , voidaan tietysti ajatella, että Z-mallissa aj-kertoimet pitävät sisällään ko. painot ja ylläoleva summa saadaan aina ykköstä pienemmäksi. Ottamalla aj-kertoimista vakiopaino $1/3$ kaikille jakaumille päästään vaadittavaan ehtoon. Liitteen Z kuviossa Z3.2 on piirretty miten tällöin sairastuneen 14 vuotiaan elossaolotodennäköisyys on asetettu laskuperusteisiin toisaalta vastaisen eläkkeen pääoma-arvoihin ja toisaalta alkaneiden eläkkeiden pääoma-arvoihin kaavojen (Z.8) ja (Z.9) mukaan ((Z.9) kukin komponentti on siis kerrottu luvulla $1/3$ suhteen suhteen z_j/z sijaan). Huono puoli tästä oletuksesta on, että kaavan (Z.7) mukaiset sairastumisintensiteetit kuviossa Z3.1 kasvavat kolminkertaisiksi.

Jos vastaisen eläkkeen puolella jakaumien paino on vakio, johtaa tämä teorian kannalta huolestuttavaan tilanteeseen, jossa etuja ostettaessa (vastaiset eläkkeen pääoma-arvot) eläkkeen päättvyys on erilainen kuin eläkettä nautittaessa käytettävä (alkaneen eläkkeen pääoma-arvot). Tietysti aina voidaan pakottaa myös vastaisiin eläkkeisiin samat painot kuin alkaneiden eläkkeiden pääoma-arvoissa kertomalla ja jakamalla suhteilla z_j/z . Tällöin (Z.6) on

$$e^{-a_4 \cdot (d-u)} \cdot \sum_{j=0}^2 \left[a_j \cdot e^{(b_j^*) \cdot (t-u)} \cdot \frac{z(t-u,0)}{z_j(t-u,0)} \right] \cdot \frac{z_j(t-u,0)}{z(t-u,0)} \cdot e^{-(c_j-b_j) \cdot u} \quad (Z.9)$$

Tätä vastaava sairastumisintensiteetti (hakasulkulauseke) on kolmas kuvaaja liitteen Z kuviossa Z3.1. Joka tapauksessa Z-mallin sairastumisintensiteetti on saatettavissa kolmen Gompertz-intensiteetin summaksi. On myös huomattava, että vakiointensiteetti a_4 sisältää - oli se sitten minkä muotoinen tahansa - sekä sairastumisintensiteetin että kuolemisintensiteetin, kuten luvuista X ja Y muistetaan.

Z.5. Z-mallin pääoma-arvot

TEL:n mukaisen vakuutuksen yleisten laskuperusteiden mukaan vastaisen eläkkeen pääoma-arvo lasketaan kaavasta

$${}_{(e)}\overline{A}_{x:w} = e^{a_4 \cdot x} \int_{x+e}^w \int_e^{t-x} e^{-\delta \cdot t} \cdot z(t,u) du dt \quad (Z.10)$$

missä x on tarkasteluikä, w on vanhuuseläkkeelle siirtymisikä. Esitys (Z.10) vastaa suoraan esitystä (Y.5) ja (X.8). Laskuperusteissa on huomioitu odotusaika e ja eläkkeen päättymisen aina vanhuuseläkeiässä w . Alkaneen eläkkeen pääoma-arvo saadaan kaavasta

$$\bar{a}_{[x]+(t-u):w}^{\overline{ii}} = \frac{1}{e^{-\delta \cdot t} \cdot z(t, t-x)} \cdot \int_t^w e^{-\delta \cdot s} z(s, s-x) ds. \quad (Z.11)$$

Mikä vastaa esitystä (X.4). Lisäksi laskuperusteissa on määritelty vuotuinen etukäteinen riskimaksu

$${}_{(e)}\bar{\Pi}_{x:w} = {}_{(e)}\bar{A}_{x:w} - e^{-(a_4+\delta) \cdot x} {}_{(e)}\bar{A}_{x+1:w}. \quad (Z.12)$$

Z.6 Työkyvyttömyyden ja työkyvyttömyyseläkkeen alkamisen erottaminen

Eläkevakuutusyhtiön kannalta aktiivin sairastuminen ei vielä ole toimenpiteitä aiheuttava tieto. Eläkkeen henkilö saa vasta odotusajan e (ja tietysti myöntämisen) jälkeen. Koska Z-mallissa on määritelty vain intensiteetit aktiivinaoloon, sairastumiseen ja työkyvyttömänäoloon, ei siitä ole suoraan saatavissa intensiteettejä aktiivinaoloon ja alle e :n mittaiseen sairastumiseen, työkyvyttömyyseläkkelle siirtymiseen tai työkyvyttömyyseläkkeen päättymiseen. Tuomikoski [1986] on selvittänyt tätä ongelmaa SHV-tutkielmassaan hyvin kattavasti. Suoraviivaisin tapa hoitaa odotusaikaan liittyvät asiat olisi siirtää tarkastelemaan työkyvyttömyyseläkkeen alkamista sairastumishetken sijasta, jolloin $e = 0$. Tämä ratkaisu ei kuitenkaan ole aivan ongelmaton. Tällöin pitäisi estimoida kullekin odotusajalle (joita TEL:n piirissä kuitenkin on useita) oma mallinsa. Samoin mikäli odotusaikaa muutetaan olisi laskuperusteet taas uusittava. Käytännössä on parempi sovittaa yksi malli, jota voi soveltaa kullekin odotusajalle, josta sairastuminen on riippumaton. Kaavat (X.8) ja (Y.5) on siis päivitettävä vastaamaan kaavaa (Z.10). Mainittakoon, että Pitaccon [1995] ratkaisu odotusaikaan liittyvään ongelmaan, johtaa täsmälleen samanlaiseen ratkaisuun.

Pääteikään liittyvät ongelmat näyttävät olevan helpommin hoidettavissa. Äärettömän sijaan astetaan pääteiksi vanhuuseläkeikä w . Tarkastellaan mitä tämä merkitsee estimoitavien jakaumien suhteen. Vaikka työkyvyttömäksi voidaan TEL:n mukaan joutua vain ikävälillä 14 - 65, ei ole syytä jakaumia estimoidessa ottaa todennäköisyysmassaa huomioon vain tältä rajoitetulta alueelta eli käyttää katkaistuja jakaumia. Sairastuminen tai paraneminen ei saisi riippua maksettavasta edusta, se voi tapahtua myös esimerkiksi iän 65 jälkeen. Ylhäältä katkaistua jakaumaa estimaattaessa saattaisi elossaolotodennäköisyys $S(t) = 1 - F(t)$ pakottua nolaksi ikäluokkaan 65 mennessä, mikä selvästi johtaisi harhaiseen tilanteeseen: kaikki henkilöt eivät sairastu varmasti tai sairaana ollessaan parane tai kuole ikään 65 mennessä. Sen sijaan alhaalta katkaiseminen ei johda tällä tavalla harhaan vaan tarkempaan malliin. Ajatellaan, että tarkasteltava perusjoukko pitää sisällään vain ne henkilöt, jotka

selviytyvät työelämään asti. Näitä henkilöitä TEL-yhtiöt vakuuttavat. Joka tapauksessa havaintoja näiltä ikäalueilta ei saada TEL-yhtiöiden kannoista. Satunnaismuunnos z on ainakin näin ollen päivitettävä seuraavasti jo sen takia että voi olla $w < 65$.

$$z(T, J) = \begin{cases} J=i & J \neq i \\ 0 & 0 & , T < s \\ e^{-\delta \cdot s} \cdot \overline{a_{T-s}} & 0 & , s < T < w \\ 0 & 0 & , T > w \end{cases}$$

Tämä satunnaismuuttuja johtaa samaan pääoma-arvoon kuin yllä kaavassa (Z.10).

Z.7 Z-malli tässä työssä

Z-mallille on TEL:n laskuperusteissa annettu erittäin vaativa tehtävä. Sen olisi kuvattava oikein tapahtumaketju, jossa vastasyntynyt elää tiettyyn ikään, jolloin hän sairastuu ja on tämän jälkeen työkyvytön pääteikänsä asti. Samalla z -funktiolla saadaan kaikki mahdolliset sairastumisiät käsitellyiksi. Vastaisia eläkkeitä varten siitä on diskontattava pois aktiivin elossaolotodennäköisyyttä, jotta saadaan tarkasteltua iäkiä $x > 0$. Intensiteettien voidaan ajatella riippuvan sairastumisiästä ja tulevasta sairausajasta. Alkaneita eläkkeitä varten Z-mallin komponenttien z_j avulla Z-mallin alkaneen eläkkeen päätyvyys sekoitetaan uudelleen, jotta malli kuvaa tarkemmin todellisten eläketapausten poistuvuutta. Vastaisten eläkkeiden puolella sekoitussuhdetta ei ole eksplisiittisesti sanottu. Z-funktio on määritelty vain vastasyntyneelle ja sillä on implisiittinen arvo sairauden kestoille, jotka ovat lyhyintä eläkkeen kestoa lyhyempiä.

Z-mallin estimointi on havaittu ongelmalliseksi. Esimerkiksi havaintoaineiston kokoamiseen on liittynyt ongelmia (Tuomikoski [1986]). Aktiivin elossaolotodennäköisyyden määrittäminen ennen sairastumista on jo teoreettisesti vaikeaa. Ei ole mahdollista selvittää jo kuolleiden joukosta niitä, jotka eläessään myöhemmin olisivat sairastuneet, mikäli "satunnaiskulku" olisi johtanut siihen, että juuri he eivät olisi kuolleet. Myös tulevaisuudessa tämä mallin osa on estimoitava muunlaisesta aineistosta. Z-mallia estimoidessa puhdasta suurimman uskottavuuden (ML) ratkaisua ei ole saavutettu. Tämä onkin vaikeaa, koska vähimmilläänkin mallissa on yhdeksän parametria. Puhdas ML-ratkaisu on vaikeasti saavutettavissa, koska mallia vaivaa paha multikollineaarisuusongelma. Kaksi sovitettavan epälineaarisen mallin selittävästä muuttujasta esiintyy mallissa kumpikin kolme kertaa eli samalle selittävälle muuttujalle haetaan kolmea eri kerrointa. Lineaaristen mallien yhteydessä tämä ongelma

lievemmissäkin muodoissa johtaa parametrien epästabiilisuuteen. Multikollineaarisuutta on yritetty lieventää valitsemalla etukäteen osa mallin parametreista (odotusajat), jolloin multikollineaarisuus lievenee ja jonkinlainen ratkaisu voidaan saavuttaa. Kuten Tuomikoski [1986] huomauttaa Z-malli on pahasti yliparametroitu.

Tässä työssä työkyvyttömyystodennäköisyyksiä ei enää yritetä sovittaa suoraan Z-malliin. Tässä työssä yritetään sovittaa mallit erikseen terveenä olemiselle, sairastumiselle ja sairastumisen päättymiselle eli kukin Z-mallin kolmesta osasta erikseen.

A. Aineiston käsittely

Elinaikamallit vaativat aineiston esikäsittelyä ennen mallin sovittamista. Tätä tehtävää selvitetään tässä luvussa.

A.1 Aineiston epäparametrinen sovittaminen

Tämän kappaleen tarkastelu on tehty Gerberiä [1980] sekä Cox, Oakesia [1984] mukaillen. Tutkitaan aluksi kuolevuutta. Tarkastellaan henkilöä i , joka täyttää iän x vuonna v . Vuoden alussa hänet otetaan mukaan tarkasteluun. Mukaanotto saattaa viivästyä esimerkiksi siksi, että henkilön työnantajan vakuutus alkaa TEL-yhtiössä vasta vuoden v aikana tai että henkilö on tullut töihin vasta ko. vuoden aikana. Merkitään aikaa vuoden alusta tarkasteluun mukaanottoon merkinnällä t_i . Vuoden lopussa hän poistuu tarkastelusta vuoden alusta laskien ajan s_i kuluttua. Henkilöllä, joka pysyy tarkastelussa koko vuoden, on selvästi $s_i - t_i = 1$. Erotusta $s_i - t_i$ kutsutaan henkilön i seurannassaoloajaksi vuonna v . Luku s_i pienenee vastaavasti kuin t_i mikäli vakuutus tai työsuhde loppuu ennen vuoden vaihdetta, mutta myös mikäli henkilö kuolee, tulee työkyvyttömäksi tai siirtyy muulle eläkkeelle. Nyt kiinnostuksen kohteena oleva tapahtuma on kuolema. Merkitään ikäluokasta x seurannassaoloaikana kuolevien henkilöiden joukkoa merkinnällä I ja siis kuolema tapahtuu henkilölle i hetkellä s_i vuonna v . Muut henkilöt - hengissä selvinneet - ovat ns. *katotapahtumia*; merkitään joukkoa J . Selvästi $0 \leq t_i \leq s_i \leq 1$. Koko kohortti-ikäluokan x seurannassaoloaika on siis, kun henkilöitä ko ikäluokassa on n kappaletta

$$E_x = \sum_{i=1}^n (s_i - t_i), \quad (\text{A.1})$$

ja kuolevien lukumäärä on D_x . Likelihood-funktioksi saadaan

$$\begin{aligned} L &= \prod_{i \in I} f(s_i | T > t_i) \cdot \prod_{i \in J} P(T > s_i | T > t_i) \\ &= \prod_{i \in I} \mu_x(s_i) \cdot P(T > s_i | T > t_i) \cdot \prod_{i \in J} P(T > s_i | T > t_i). \end{aligned} \quad (\text{A.2})$$

Oletetaan edelleen, että kuolevuus on koko vuoden v iässä x vakio; merkitään $\mu_x(s_i) = \mu_x$. Tällöin likelihood-funktio saadaan muotoon

$$L = (\mu_x)^{D_x} \cdot e^{-\mu_x \cdot E_x}, \text{ jonka maksimi saavutetaan, kun } \hat{\mu}_x = \frac{D_x}{E_x}.$$

Näin ollen ikäluokan x ML-estimaattori kuolemistodennäköisyydelle on siis $\hat{q}_x = 1 - e^{-\hat{\mu}_x} = 1 - e^{-(D_x/E_x)}$. Elossaolotodennäköisyys iässä x vuoden v loppuun asti on

$$1 - \hat{F}(x + 1/2) = \hat{S}(x + 1/2) = \prod_{k=0}^x (1 - q_k) = \prod_{k=0}^x e^{-(\frac{D_k}{E_k})}. \quad (\text{A.3}) \text{ ja}$$

$$\hat{E}(x) = \sum_{x=0}^{\infty} \hat{q}_x$$

Tämä on vakuutusmatematiikan piirissä yleisin tapa laskea havaitut elossaolotodennäköisyydet ja intensiteetit. Tätä yllä kuvattua menetelmää ei ole tietysti pakko käyttää vain vuosittaiseen tarkasteluun, vaan epäparametriset estimaatit voidaan sovittaa myös koko elinajan tai työssäoloajan kattavalle tarkasteluvälille. Esityksen (A.1) mukainen likelihoodfunktio on esitys yleisemmästä puuttuvien havaintojen käsittelystä ML-estimoinnissa.

Jaetaan vielä kuoleminen kahteen (kilpailevaan) syyhyn i ja j , jolloin $D_x = D_{i,x} + D_{j,x}$. Nyt likelihoodfunktio saa muodon

$$L = \prod_{k=i}^j (\mu_{k,x})^{D_{k,x}} \cdot e^{-\mu_x \cdot E_x},$$

kun taas oletetaan, että intensiteetit ovat vakioita koko vuoden v . Likelihoodfunktion maksimi saavutetaan, kun $\hat{\mu}_{k,x} = \frac{D_{k,x}}{E_x}$, $k = i, j$.

Huomautettakoon yhdestä approksimaatiosta, joka usein tehdään, kun havaittua elossaolotodennäköisyyttä arvioidaan ja kun intensiteetit ovat pienet:

$$e^{-x} = 1 - x + o(x^2) \Rightarrow \hat{q}_x = 1 - e^{-\frac{D_x}{E_x}} \approx \frac{D_x}{E_x}.$$

Tässä luvussa Gerberistä [1980] poiketen tarkasteltiin kohortti-ikäluokkaa x verrattuna perinteiseen ikäluokkaan x . Kohortti-ikäluokka x saavuttaa iän x vuoden v aikana. Perinteinen ikäluokka x tarkoittaa henkilöitä, jotka ovat jo täyttäneet iän x eli he kuuluvat ikäluokkaan x vuosien v ja $v+1$ aikana (yhden vuoden ajan). Kohortti-ikäluokka on vuoden v aikana keskimäärin iässä x ja perinteinen ikäluokka x saadaan, kun ikää arvioidaan katkaisemalla henkilöiden ikä kokonaisluvuksi. TEL:n laskuperusteissa tarkastellaan vakuutettuja siten, että vuoden v aikana henkilö on keskimäärin iässä x hetkellä 1.7.v ja

iässä $x+\frac{1}{2}$ hetkellä 31.12.v., joten siihen sopii kumpi ikäluokkatarkastelu vain. Aineiston hankinnan kannalta kohortti-ikäluokkatarkastelu on kuitenkin selkeämpi.

Joissain tilanteissa esityksen (A.1) mukainen seurannassaoloaika voidaan joutua tai halutaan arvioida (ks. Elandt-Johnson, Johnson [1980]). Näin on esimerkiksi tilanteessa, jossa on käytössä vain lukumäärätietoja tai mikäli aineisto on niin laaja, että arviointimenettelyn uskotaan riittävän.

Seurannassaoloaikaa voidaan arvioida seuraavasti. Olkoon henkilöiden luku kohortti-ikäluokassa x hetkellä 31.12.v määrä N_x . Saman kohortti-ikäluokan lukumäärä edellisvuonna 31.12.v-1 (silloin iässä $x-1$) on N_{x-1} . Jos oletetaan, että seurannasta katoaa henkilöitä vain kuoleman kautta ja että kuolemistapaukset sattuvat tasaisesti vuoden v aikana, tällöin saadaan

$$E_x = N_{x-1} - \frac{1}{2} \cdot D_x.$$

Eläkeyhtiöiden kannassa vakuutettujen lukumäärässä tapahtuu myös muita muutoksia kuin kuolemia kuten yllä selostettiin. Työnantajat ottavat uusia työntekijöitä palvelukseen tai irtisanovat heitä tai työntekijät vaihtavat työpaikaa, mahdollisesti myös ei TEL:n alaisuudessa vakuutetulle alalle. Tällöin muutos lukumäärästä N_{x-1} lukumäärään N_x pitää sisällään paitsi kuolemistapahtumia myös seurannasta poistuvia ja seurantaan tulevia. Jos oletetaan, että nämä muut henkilövaihdokset tapahtuvat tasaisesti vuoden v aikana, niin keskimääräinen seurannassaoloaika on

$$E_x = N_{x-1} - \frac{1}{2} \cdot D_x + \frac{1}{2} \cdot [N_x - (N_{x-1} - D_x)] = \frac{1}{2} \cdot [N_x + N_{x-1}]. \quad (A.4)$$

Saman kohortti-ikäryhmän keskiarvo vuoden v aikana arvioi näillä oletuksilla seurannassaoloaikaa E_x .

Kuten Tuomikoski [1986] huomauttaa esityksen (A.3) mukainen elossaolotodennäköisyys voidaan laskea kahdella tavalla.

Tavassa A seurataan kutakin kohortti-ikäluokkaa erikseen. Tällöin kuolevuus pitäisi laskea kullekin syntymävuosiluokalle kussakin iässä erikseen ja aineiston olisi katettava ainakin koko TEL:n tämänhetkinen voimassaoloaika eli vuodet 1962 - 1996. On tietysti selvää, että tämä tapa vaatii paljon aineistoa.

Tavassa B lasketaan esitys (A.3) viimeisempien vuosien perusteella. Tällöin esimerkiksi 50 vuoden ikäisen elossaolotodennäköisyyteen liitetään riski kultakin hänen aiemmalta iältään niin, että nyt esimerkiksi 30 vuoden ikäisen henkilön kuolemistodennäköisyys vuonna 1996 on sama kuin tämän 50-ikäisen ollessa 30-vuotias vuonna 1976. Itseasiassa tavassa B

ei sallita kuolevuuden muutoksia. On hyvinkin mahdollista, että nyt 50-vuotias on "kovemman koulun läpikäynyt" 30-vuotiaana kuin nyt 30-vuotiaat. 50-vuotiaat voivat olla valikoidumpaa joukkoa kuin mitä tällä hetkellä 30-vuotiaan riski ilmoittaa ja tällä voi olla vaikutusta hänen selviämiseen ikään 51 vuotta. Tietysti on muistettava, että pääoma-arvoja käytetään "tulevaisuuden ennustamiseen". 50-vuotiaalle henkilölle pääoma-arvo lasketaan siten, että häntä vanhempien henkilöiden kuolintodennäköisyydet otetaan ko henkilön tulevaisuuden riskiksi. Estimointitapojen A ja B vertailu olisi järkevää ja voi tietysti olla ettei tapojen välillä ole eroa.

Elossaolotodennäköisyyden varianssin estimaattorina voidaan käyttää Greenwoodin [1926] kaavaa

$$Var[\hat{S}(x + \frac{1}{2})] = [\hat{S}(x + \frac{1}{2})]^2 \sum_{k=1}^x \frac{D_k}{N_k(N_k - D_k)}.$$

A.2 Aineiston parametrinen sovittaminen

Tässä luvussa on käytetty lähteenä Cox, Oakes [1984] ja Kalbfleish, Prentice [1980] sekä Elandt-Johnson, Johnson [1980]. Käytettävien termien suomennoksessa on käytetty lähteenä Hakulista [1983].

Mikäli elinaikamalleissa olisi käytössä $n:n$ havainnon riippumaton otos, joista kaikista tiedettäisiin seurannassaoloajan päättymisen tutkittavaan tapahtumaan, olisi likelihoodfunktio yleisen esityksen mukainen:

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta; z_i) = \prod_{i=1}^n \mu(t_i; \theta; z_i) \cdot S(t_i; \theta; z_i), \quad (\text{A.5})$$

jossa t_i on havaittu eliaika (ei siis vain tiettyä vuotta v), θ on esitimoitava parametrivektori ja z_i on (vielä toistaiseksi tarpeeton) ns. selittävien muuttujien vektori. Käytännössä kaikki seurannassaolevat henkilöt eivät kuole seurannan päättämishetkeen mennessä tai osa siirtyy muuten pois seurannasta esimerkiksi vanhuuseläkkeelle siirtymisen takia. Näin ollen näiden seurannasta poistuneiden kuolinaika jää tuntemattomaksi eikä sitä voida saada selville ja ainoa mitä heistä tiedetään on, että he olivat elossa seurannasta poistuessaan. Nyt otokseen on kerättävä uusi tieto δ_i siitä onko elinaika t_i kuolinaika ($\delta_i = 1$) vai seurannasta poistuminen ns. katotapahtumaan ($\delta_i = 0$). Nyt likelihoodfunktio saa muodon

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n f(t_i; \theta; z_i)^{\delta_i} \cdot S(t_i; \theta; z_i)^{1-\delta_i} \\ &= \prod_{i=1}^n \mu(t_i; \theta; z_i)^{\delta_i} \cdot S(t_i; \theta; z_i), \end{aligned} \quad (\text{A.6})$$

jossa jälkimmäinen termi seuraa esityksestä (X.2) ja vastaa esitystä (A.1).

Mikäli mallien estimointi tehdään kahdessa vaiheessa, niin että ensin lasketaan ns. epäparametriset luvun A.1 mukaiset estimaatit ja tämän jälkeen lähdetään sovittamaan varsinaista parametrista mallia, on likelihoodfunktiona käytettävä esityksen (A.5) mukaista muotoa. Katotapahtumat on jo huomioitu epäparametrisessä estimoinnissa. Jos taas aineistoa ei ole koottu luvun A.1 mukaiseen muotoon ja aineistona käytetään suoraan havaittuja elin- tai katoaikoja on esitystä (A.6) käytettävä.

Mikäli saatavilla oleva aineisto on sellainen, että siinä on vain tietyn hetken τ_j jälkeen voimassa olevat ja voimaan tulevat seuranta-ajat, on likelihoodfunktio muutettava. Tällöin kaikki ennen aikaa τ_j päättyneet havainnot puuttuvat ja likelihoodfunktioon

vaikuttavaa jakaumaa on pidettävä katkaistuna. Tällöin (A.6) saa muodon (ja (A.5):ttä on korjattava vastaavasti), joka vastaa ehdollista todennäköisyyttä

$$\begin{aligned}
 L(\theta) &= \prod_{i=1}^n \frac{f(t_i; \theta; z_i)^{\delta_i}}{S(\tau_i; \theta; z_i)^{\delta_i}} \cdot \frac{S(t_i; \theta; z_i)^{1-\delta_i}}{S(\tau_i; \theta; z_i)^{1-\delta_i}} \\
 &= \prod_{i=1}^n \mu(t_i; \theta; z_i)^{\delta_i} \cdot \frac{S(t_i; \theta; z_i)}{S(\tau_i; \theta; z_i)} .
 \end{aligned}
 \tag{A.7}$$

Likelihoodfunktion maksimointi vaatii omaa tekniikkaa. Tietysti paras ratkaisu olisi, jos löydettäisiin ns. analyttinen ratkaisu. Usein käytännössä maksimointi joudutaan tekemään iteratiivisesti. Tähän liittyvää tekniikkaa on selostettu kattavasti Tuomikosken [1986] tutkielmassa. Usein likelihoodfunktioista otetaan logaritmi, jolloin tulon sijasta tarkastellaan summaa ja iteratiiviset menetelmät tuottavat joskus varmemmin ratkaisun, jos etsitään maksimin sijasta minimiä.

Mallien sovittamista tullaan kokeilemaan myös ns. regressiomenetelmillä, joita on selostettu useissa oppikirjoissa mm. Seber [1977], yleistettyjen lineaaristen mallien osalta kirja McCullagh, Nelder [1989] on perusteos. Elinaikamalleihin sovellettuna regressiomenetelmiä on selostettu jonkin verran kirjassa Kalbfleish, Prentice [1980].

A.3 Aineiston käyttö tässä työssä

Tätä työtä varten tarvittiin aineistoa eläkeyhtiöistä. Työkyvyttömyysilmiötä TEL:n piirissä selvittelee ns. työkyvyttömyysjaos, joka on TELA:n alainen toimikunta. Toimikunnalle eläkeyhtiöt toimittavat vuosittain tiedot uusista ja päättyneistä työkyvyttömyyseläketapauksista. Toimikunnan hallussa onkin kattavin aineisto työkyvyttömyysilmiöstä, jonka vain voi TEL:n piiristä saada - sillä on aineistoa aina 1960-luvulta asti. Työkyvyttömyysjaokselle ei kuitenkaan mene tietoa yllä kuvatusta seurannassaoloajasta. Nämä tiedot täytyy hankkia muista lähteistä eli lähinnä eläkelaitosten työsuuhdekannoista. Myöskään eläkelaitosten vuosittain tekemistä riskiin liittyvistä tutkimuksista ei ollut tässä yhteydessä suoraa hyötyä.

En kuitenkaan onnistunut saamaan tätä työkyvyttömyysjaoksen kattavaa aineistoa käyttööni, joten jouduin tyytymään Eläke-Varmasta saatavaan aineistoon. Seuraavassa on selostettu miten ja millainen aineisto hankittiin ja mitä oletuksia jouduttiin tekemään.

A.3.1 Kuolevuus ja sairastuvuus

Kuolevuus ennen sairastumista on - kuten aiemmin todettiin - vaikea tutkittava. Jo kuolleiden joukosta ei voida etsiä niitä, jotka mahdollisesti eläessään olisivat sairastuneet niin vakavasti, että olisivat joutuneet työkyvyttömyyseläkkeelle, mikäli olisivat eläneet. Siksi aineistona täytyy käyttää TEL-yhtiöiden yleistä kuolevuutta. Koska TEL-vakuutuksessa vanhuuseläkeliiikkeessä on jo käytössä kuolevuusfunktio, joka uusittiin viimeksi vuoden 1997 alussa, ei katsottu järkeväksi sovittaa uutta mallia kuolevuudelle. Vain uuden mallin sopiminen havaintoaineistoon tarkastetaan elossaolotodennäköisyyden avulla. Kovin järkevää ei ole, jos saman vakuutustoiminnan sisällä on useita kuolevuusfunktioita.

Aineisto saadaan työsuuhdekannoista. Kuolevuuden osalta päätettiin ensin laskea epäparametriset estimaatit tapaa A käyttäen ja käyttää kaavaa (A.1) seurannassaoloajan laskemiseksi. Tavan A mukaiset estimaatit lasketaan ja tätä verrataan kuolevuusperusteen mukaiseen arvoon. Luokittelu tehdään erikseen molemmille sukupuolille sekä siis erikseen kullekin syntymävuosi-ikäluokalle.

Eläke-Varman työsuuhdekannasta pyrittiin poimimaan kaikki aikavälillä 1.1.1962 - 31.12.1996 voimassa olleet todelliset työsuhteet. Mukaan otettiin myös luottamustoimet, mutta esimerkiksi peruutetut ja Valtionkonttoriin siirretyt jaksot pudotettiin pois. Koska vakuutetulla saattaa olla työsuhteen rinnalla myös samaan aikaan rinnakkaisia eläkejaksoja, vähennettiin kunkin henkilön osalta nämä aikajaksot seurannassaoloajasta pois. Mikäli

henkilöllä on useita työjaksoja yhtä aikaa otettiin mukaan seurannassaoloaikaan vain yksi aikaväli samaan aikaan palvelusta työajasta. Asevelvollisuus aika sekä äitiysloma- ja lapsenhoitoaika sen sijaan jätettiin seurannassaoloaikaan yksinkertaisesti siksi, että olisi todennäköistä, että juuri kyseessä oleva eläkelaitos myöntäisi niin käydessä tällaiselle henkilölle työkyvyttömyyseläkkeen. Yksilöllisen varhaiseläkkeen levätessä tämä aika vähennettiin.

Eläke-Varman työsuuhdekannassa on yli 4 miljoonaa työjaksoa. Koska käytössä oli vain rajallisesti atk-resursseja, jouduttiin henkilöt rajaamaan niin, että tarkasteluun otettiin mukaan vain kunkin kuun kolmantenatoista päivänä syntyneet. Tällaisia jaksoja löytyi yhteensä 130 840 jaksoa, jossa oli 60 042 henkilöä mukana. Näiden henkilöiden joukosta etsittiin sitten myös kuolemaan päättäneet jaksot, joissa henkilöitä on yhteensä 438 kappaletta. Sama henkilö saattoi kuollessaan olla vakuutettu useammassakin työjaksossa; vain yksi huomioitiin.

Sairastuvuuden laskemiseksi ei työsuuhdekannasta ollut apua. Koska eläkkeiden maksurekisteri on nimensä mukaisesti vain eläkkeiden maksamista varten ja koska atk-järjestelmä oli vasta uudistettu, ei Eläke-Varmasta ollut saatavilla kattavaa kohortti-ikäluokka-aineistoa. Eläkkeiden maksurekisteri puhdistetaan aika-ajoin ns. turhista jo päättäneistä tapauksista. Sairastuvuuden laskemiseksi on tyydyttävä tässä harjoitustyössä tarkastelemaan tavan B mukaan laskettuja intensiteettejä ja jakaumia. Myös Eläke-Varman vuosilaskentajärjestelmä on vast'ikään uusittu ja sieltä oli saatavilla vuosien 1994 - 1996 vuosilaskennassa olleet uudet työkyvyttömyystapaukset. Tarkasteluun päätettiin ottaa vuosien 1994 ja 1995 aikana sattuneet eläketapahtumapäivät, jotka ovat olleet Eläke-Varman vuosilaskennoissa vuosina 1994 - 1996. Eläketapahtuma tyypillisesti selviää noin yhdeksän kuukautta sairastumispäivän jälkeen, joten vuoden 1996 vuosilaskennassa on ollut mukana vielä vuoden 1994 aikana tapahtuneita sairastumistapauksia.

Seurannassaoloaika saadaan yllä työsuuhderekisteristä saaduista luvuista, mutta vain käyttämällä vuosien 1994 ja 1995 ikäluokkakohdaista seurannassaoloaika. Vuodet yhdistettiin siten, että ikänä käytettiin ikää kyseisenä vuonna eli esimerkiksi ikäluokkaan 60 yhdistettiin vuonna 1934 syntyneet eläketapahtumavuoden ollessa 1994 ja vuonna 1935 syntyneet eläketapahtumavuoden ollessa 1995. Mukaan otettiin vain ns. oikeat työkyvyttömyyseläkkeet sekä kuntoutustuet (joista on siis peritty pääoma-arvo). Ns. yksilölliset varhaiseläkkeet jätettiin pois. Kullekin ikäluokalle voitiin näin laskea intensiteetti, mutta kohorttitarkastelua ei voitu tehdä. Kun työsuuhderekisteristä oli poimittu vain 13:nä päivänä syntyneet, oli niin tehtävä myös sairastuneille. Otokseen tuli yhteensä 117 sairastumistapauksta. Ainoa lisämuuttuja, joka saatiin mukaan, oli sukupuoli.

Aineiston vähyiden takia tässä työssä rajoitutaan ensin laskemaan ns. epäparametriset ML-estimaatit tapaa B käyttäen ja sitten sovitetaan tähän 'aineistoon' jakauma.

A.3.2 Työkyvyttömyyseläkeläisten kuolevuus ja parantuvuus

Koska työkyvyttömyyseläkeläisen kuolevuuteen tai paranemiseen liittyviä tekijöitä on paljon, hankittiin tästä osasta mallitettavaa ilmiötä toisenlainen aineisto. Taaskin käytössä on vain Eläke-Varman vuosilaskennoissa mukana olleet työkyvyttömyyseläketapaukset, joista varhaisimpien eläketapahtumavuodet sattuvat 1960-luvun alkuun. Tarkasteltava otos käsittää vuosilaskennoissa 1994 - 1996 mukana olleet tapaukset, kuitenkin niin, että sama eläkejakso tulee tarkasteluun vain kerran, vaikka se olisi ollut kaikissa vuosilaskennoissa mukana ja yksi henkilö tulee mukaan vain kerran samasta eläkkeelläoloajasta. Ns. määräaikaisten työkyvyttömyyseläkkeiden osalta tarkastelu tehtiin niin, että mikäli eläkejakso oli päättynyt ennen 1.1.1997, on kyseessä oleva henkilö parantunut ja eläkkeen päättyessä myöhemmin eläkejakso oli jatkuva eli katotapaus. Taaskin yksilölliset varhaiseläkkeet jätettiin pois, kuten myös kuntoutusrahat - sen sijaan kuntoutustuet otettiin mukaan.

Eläke-Varman kannasta löytyi yhteensä 27 828 eri työkyvyttömyystapausta, joista aikavälillä kuolemaan päättyi 1327 ja joista parantui 540. Toisin sanoen ns katotapauksia tässä aineistossa tuli olemaan yli 93 prosenttia koko aineistosta. Eläketapauksista onnistuttiin saamaan seuraavat tiedot: syntymäaika, sukupuoli, eläketapahtumapäivä, eläkelaji sekä eläkkeen päättymispäivä ja -syy. Malli sovitetaan suoraan esityksen (A.6) mukaisen likelihoodtarkastelun avulla.

A.4 Työkyvyttömyyteen liittyvät taustatekijät muissa tutkimuksissa

Työkyvyttömyysriskiä ovat tutkineet viime vuosina mm. Hännikäinen [1988] ja Hytti [1993]. Työkyvyttömyyden alkavuuden selviksi vaikuttaviksi tekijöiksi on havaittu ikä ja sukupuoli. Lisäksi on tutkittu sosiaaliryhmää (akselilla työntekijä/toimihenkilö), jolla on havaittu olevan vaikutusta alkavuuden tasoon. Myös toimialaa on tutkittu, tosin sukupuoli vaikuttaa vieläkin voimakkaasti henkilön valitsemaan toimialaan ja myös sitä kautta korreloi sairastuvuusriskin kanssa; yhteisvaikutusta ei ole tutkittu. Hytti on tutkinut myös sairastumissyitä ja havainnut, että tässäkin asiassa on olemassa vielä jako miesten ja naisten sairauksien välillä. Muita korreloivia tekijöitä voi olla lisäksi suhdanteet, työttömyysaste, tulotaso, perhesuhteet, korvaustaso sekä työkyvyttömyysmääritelmän muutokset. Hännikäinen on tutkinut myös työkyvyttömyyden päättävyyttä ja havainnut taustatekijöiksi mm. alkuiän, menneen keston, sukupuolen ja sairastumissyyn. Hytin tutkimuksessa löydettiin 0,4 % sellaisia työkyvyttömyystapauksia, joissa sairastunut henkilö kävi välillä töissä ja sairastui uudelleen, joten ilmeisesti luvun Y mukaiselle tilaluokkamallin approksimaatiolle ei ole tarvetta. Määräaikaisina jatkuvat eläkkeet oli yhdistetty yhdeksi tapaukseksi.

S. Sovitettavat mallit

S.1 Jakauman sovittaminen

Kuten luvussa X todettiin, intensiteettejä voidaan tietyin ehdoin käyttää määrittelemään jakauma. Suoraviivaisin estimointitapa olisi siis sovittaa havaittujen intensiteettien avulla (luku A.1) tiettyjä teoreettisia jakaumia. Elinaikamallien kirjallisuudessa on käytetty seuraavia jakaumia ($t \in (0, \infty)$)

Exponenttijakauma $\mu(t) = \mu, \quad \mu > 0$
 $f(t) = \mu \cdot e^{-\mu t}$
 $S(t) = e^{-\mu t}$

Weibull-jakauma $\mu(t) = \alpha \cdot \lambda \cdot (\lambda \cdot t)^{\alpha-1}, \quad \alpha, \lambda > 0$
 $f(t) = \alpha \cdot \lambda \cdot (\lambda \cdot t)^{\alpha-1} \cdot e^{-(\lambda \cdot t)^\alpha}$
 $S(t) = e^{-(\lambda \cdot t)^\alpha}$

(S.1)

Log-logistinen jakauma $\mu(t) = \frac{\alpha \cdot \lambda \cdot (\lambda \cdot t)^{\alpha-1}}{1 + (\lambda \cdot t)^\alpha}, \quad \alpha, \lambda > 0$
 $f(t) = \frac{\alpha \cdot \lambda \cdot (\lambda \cdot t)^{\alpha-1}}{[1 + (\lambda \cdot t)^\alpha]^2},$
 $S(t) = \frac{1}{1 + (\lambda \cdot t)^\alpha} = 1 - F(t) = \frac{1}{(\lambda \cdot t)^\alpha} \cdot F(t)$

Gompertz-Makeham $\mu(t) = a + \mu \cdot e^{bt}, \quad \mu > 0$
 $f(t) = (a + \mu \cdot e^{bt}) \cdot e^{-\left(a + \frac{\mu(e^{bt}-1)}{b}\right)}$
 $S(t) = e^{-\left(a + \frac{\mu(e^{bt}-1)}{b}\right)}$

Myös muita jakaumia on käytetty mm. lognormaali-, invnormaali-, gamma- ja F-jakaumaa (ks. esim. Manton etc. [1981], Bartholomew [1982]). Näiden käyttökelpoisuutta rajoittaa kuitenkin niiden monimutkaisuus käytännöllisessä työssä.

Yksi mahdollisuus on myös käyttää jakaumien sekoitusta. Esimerkiksi kahden normaalijakauman sekoitetta voidaan kokeilla, jolloin

$$\phi(t, \mu_1, \sigma_1, \mu_2, \sigma_2, p) = p \cdot \phi_1(t, \mu_1, \sigma_1) + (1-p) \cdot \phi_2(t, \mu_2, \sigma_2).$$

Tällöin painoparametri p on kuitenkin vakio, eikä riipu selittävistä tekijöistä.

S.2 Suhteellisen vaaran malli

S.2.1 Selittävien tekijöiden liittäminen intensiteetteihin

Ns. Coxin [1972] (ks. myös mm. Breslow [1972]) *suhteellisen vaaran malli* antaa erinomaiset edellytykset liittää intensiteetteihin selittäviä tekijöitä. Coxin mallin mukaan

$$\mu(t|y) = \mu_0(t) \cdot e^{\beta^T \cdot y}, \quad (\text{S.2})$$

jossa y on selittävien muuttujien vektori, joka voi olla paitsi diskreetti myös jatkuva ja β niihin liittyvä estimoitava parametrivektori sekä $\mu_0(t)$ jollakin tavalla estimoitu perusjoukon intensiteetti esim. luvun S.1 mukainen. On huomattava, että kun y saa arvokseen 0, on $\mu_0(t) = \mu(t)$ eli selittävät muuttujat vaikuttavat *suhteessa* johonkin valittuun perusjoukkoon (joilla $y=0$). Itse asiassa Coxin suhteellisen vaaran mallissa esityksen (S.2) sijaan yritetään yleensä sovittaa mallia

$$\log\left(\frac{\mu(t|y)}{\mu_0(t)}\right) = \beta^T \cdot y, \quad (\text{S.3})$$

joka on estimoitavissa esimerkiksi regressiomenetelmillä. Coxin mallissa on nähty suurta vaivaa, ettei tarvitsisi estimoida parametrisesti intensiteettejä $\mu_0(t)$ ja yleensä intensiteetteinä käytetään epäparametrisesti sovitettuja estimaatteja ns. Kaplan-Meier-estimaatteja [1958]. Toisin sanoen Coxin malli vastaa parhaiten kysymykseen paljonko vaara lisääntyy/vähenee esimerkiksi eri ryhmissä tai annettaessa tietty määrä jotain lääkettä. Coxin malli ei suoraan sovi tämän harjoitustyön tarkoitukseen vaikkakin antaa perustellun pohjan selittävien muuttujien käytölle: sovitetaan ensin jakauma tietylle perusjoukolle ja tutkitaan sitten miten muiden ryhmien vaara lisääntyy tai vähenee (multiplikatiivisesti) suhteessa tähän perusjoukkoon. Huomautettakoon vielä lopuksi, että luvun A.2 intensiteettiesitys on sama kuin (S.2). Siinä on lisänä vain parametrivektori θ_1 , joka pitää sisällään paitsi β -vektorin myös perusjoukon sovitettavan jakauman parametrit.

S.2.2 Regressiomenetelmillä sovitettavat mallit

EkspONENTTijakaumaan selittävien tekijöiden liittäminen Coxin mallin mukaan on suoraviivaista. Tällöin

$$\mu(t|y) = \mu \cdot e^{\beta^T \cdot y}, \quad \text{ja} \quad \log(-\log(S(t|y))) = \log(\mu) + \log(t) + \beta^T \cdot y. \quad (\text{S.4})$$

Aivan vastaavasti Weibull-jakaumalle saadaan

$$\mu(t|y) = \alpha \cdot \lambda \cdot (\lambda \cdot t)^{\alpha-1} \cdot e^{\beta^T \cdot y}, \quad \log(-\log(S(t|y))) = \alpha \cdot \log(\lambda) + \alpha \cdot \log(t) + \beta^T \cdot y, \quad (\text{S.5})$$

joten eksponenttijakaumaa ei kannata kokeilla vaan se tulee valituksi vain, jos Weibull-jakaumalle $\alpha=1$. Yleistettyjen mallien piirissä ko. elossaolotodennäköisyydelle tehty muunnos tunnetaan ns. komplementaarisena log-log muunnoksena.

Log-logistiselle jakaumalle saadaan toisella tavalla

$$\mu(t|y) = \frac{\alpha \cdot \lambda \cdot (\lambda \cdot t \cdot e^{\beta^T \cdot y})^{\alpha-1}}{1 + (\lambda \cdot t \cdot e^{\beta^T \cdot y})^\alpha}, \quad \log\left(\frac{F(t|y)}{S(t|y)}\right) = \alpha \cdot \log(\lambda) + \alpha \cdot \log(t) + \alpha \cdot \beta^T \cdot y \quad (\text{S.6})$$

mikä vastaa yleistettyjen lineaaristen mallien piirissä ns. odd-ratio muunnosta ja malli on seuraavassa luvussa esiteltävä kiihtyneen tapahtuma-ajan malli.

Mallien (S.5) ja (S.6) toimivuutta voidaan 'testata' piirtämällä hajontakuviioon vastakkain $x = \log(t)$ ja $y = \log(-\log(\hat{S}(t)))$ sekä toisaalta $x = \log(t)$ ja $y = \log\left(\frac{\hat{F}(t)}{\hat{S}(t)}\right)$.

S.3 Kiihtyneen tapahtuma-ajan malli

Tässä kappaleessa *kiihtyneen tapahtuma-ajan malli* on käsitelty Klein, Moeschbergerin [1997] mukaan. Mallin mukaan yksilön, jonka selittävien tekijöiden arvot ovat y , hetkellä $T=t$ elossaolotodennäköisyys on sama kuin yleinen elossaolotodennäköisyys hetkellä $t \cdot \exp(\beta^T \cdot y)$. Toisin sanoen

$$S(t|y) = S(t \cdot e^{\beta^T \cdot y}) \text{ kaikilla } t. \quad (\text{S.7})$$

Termiä $\exp(\beta^T \cdot y)$ kutsutaan kiihtyvyytermiksi. Intensiteetiksi saadaan tällöin

$$\mu(t|y) = e^{\beta^T \cdot y} \cdot \mu(t \cdot e^{\beta^T \cdot y}) \text{ kaikilla } t. \quad (\text{S.8})$$

Selittäviä tekijöitä voidaan liittää mallin myös seuraavasti

$$s = \log(t) = \mu + \gamma^T \cdot y + \sigma \cdot W. \quad (\text{S.9})$$

Tässä γ kertoimet ovat regressiokertoimia ja W on virhejakauma, vastaavasti kuin regressioanalyysissä. Nyt yllä kuvattu log-ajan malli on kiihtyneen tapahtuma-ajan malli, jos $S(t)$ on elossaolotodennäköisyys satunnaismuuttujalle $e^{\mu + \sigma \cdot W}$ ja $\beta = -\gamma$.

Virhejakaumana W on yleisesti käytetty ns. extreme value jakaumaa, joka johtaa Weibull elossaolotodennäköisyyteen, logistista jakaumaa, joka johtaa log-logistiseen elossaolotodennäköisyyteen sekä normaalijakaumaa, joka johtaa lognormaalijakauman mukaiseen elossaolotodennäköisyyteen. Weibull-jakautunut elossolotodennäköisyys on myös Coxin suhteellisen vaaran mallin mukainen, jolloin Coxin mallin parametriesitykselle pätee

$$h(t|Y=y) = (\alpha \cdot \lambda \cdot (\lambda \cdot t)^{\alpha-1}) \cdot e^{\beta_T \cdot y}, \text{ jossa } \alpha = \frac{1}{\sigma}, \quad \lambda = e^{-\mu/\sigma}, \quad \beta_j = -\gamma_j / \sigma. \quad (\text{S.10})$$

Mainittakoon lopuksi, että Gompertz-Makeham -jakauma ei ole suoraan linearisoitavissa muunnosten avulla, mutta TEL:n vanhuuseläkeliiikkeessä Gompertz-kuolevuutta ja selittäviä muuttujia on käytetty vastaavalla tavalla, tosin ajan t muunnoksena - ns. ikäsiirroksi muunnettuna

$$\mu(t|y) = \mu \cdot e^{b \cdot t} \cdot e^{\beta_T \cdot y}, \quad \text{jolloin } S(t|y) = e^{-\left(\frac{\mu \cdot e^{b \cdot t} \cdot (e^{\beta_T \cdot y} - 1)}{b}\right)}. \quad (\text{S.11})$$

S.4. Muita mallittamistapoja

Elosaolotodennäköisyyden mallittamista on useassa lähteessä tehty siten, että aika-akseli on luokiteltu ja näin saatuun luokiteltuun aineistoon on käytetty yleistettyjen lineaaristen mallien teoriaa. Renshaw, Haberman [1995] jakoivat työkyvyttömyyden päättymisen tutkimiseksi eri iät ja työkyvyttömyyden kestot luokkiin. He olettivat, että kussakin luokassa havattu parantuneiden lukumäärä on poisson-jakautunut parametrinä tulo seurannassaoloajasta ja parantumisintensiteetistä. Ns. linkkifunktio on log-muunnos. He sovittivat ensin kullekin luokalle omat päävaikutusparametrit ja havaittuaan, että ko parametrit käyttäytyvät lineaarisesti, sovittivat mallin, jossa sekä ikä että kesto on jatkuva, jolloin parametrien määrä väheni merkittävästi. Tämä on erittäin mielenkiintoinen tapa sovittaa tarvittava malli; huomautettakoon, että tällöin sovitetaan esityksen (X.5) mukaista ehdollista jakaumaa. Myös Laird, Olivier [1981] käyttävät vastaavalla tavalla ajan luokittelua ja poisson-jakaumaoletusta elossaolomallin sovittamiseen. Hakulinen [1986] käyttää suhteellista elossaolotodennäköisyyttä syöpäaineistoon mallittaessaan myös diskretisoitua aikaa. Hän olettaa, että aikavälin alussa oleva joukko, joilla on riski kuolla ko aikavälillä, on kiinteä. Tällöin kuolleiden määrä luokassa noudattaa binomijakaumaa. Tarvittava linkkimuunnos on komplementaarinen log-log-muunnos. Aranda-Ordaz [1983] ja Tibshirani, Ciampi [1983] tutkivat multiplikatiivisen ja additiivisen vaaran lisääntymisen eroja. Tietyllä tavalla luonnollisempi vaaran lisääntymiskeino on

additiivinen lisääntyminen, joka on perusteltavissa esimerkiksi kilpailevien kuolinsyiden teorian kautta (ks. esim luku X.4). Aranda-Ordaz esitti muunnoksen, jolla voidaan tutkia kumpi vaaran lisääntymiskeino sopii aineistoon paremmin. Itseasiassa muunnos on sellainen, että vaaran lisääntymiskeino voi olla myös jotain niiden väliltä. Tibshirani, Ciampi esittivät menetelmän, jolla muunnos voidaan tehdä yleistettyjä lineaarisia malleja sovittamalla. Jasiulewicz [1997] on sovittanut suoraan eri jakaumien sekoituksia havaittuihin Puolan ja USA:n elossaolotodennäköisyyksiin. Työkyvyttömyysmalleja Hollannissa on käsitellyt Gregorius [1993], Hertzman [1993], Sansom, Waters [1988], Scott [1988] on käsitellyt vastaavia malleja Isossa Britaniassa sekä Segerer [1993] Saksan, Itävallan sekä Sveitsin malleja. Puzey [1995] on käsitellyt q_x esitimaattoria. Portnoy [1997] on selvittänyt kokonaan toisenlaista kuolevuustaulujen estimointimenetelmää.

S.5. Mallin valintakriteeri

Mallit tullaan - tietysti - valitsemaan pääasiassa sen perusteella miten ne sopivat aineistoon. Keinoina tämän toteamiseksi käytetään kuvia sekä osaksi eräitä testejä.

Koska malleja tarvitaan elossaolotodennäköisyyden $S(t)$ estimointiin, on valittavan mallin sovittava tähän havaittavaan elossaolotodennäköisyyteen. Siksi kuviota, jossa on sovitettu elossaolokäyrä ja sovitettu malli, tullaan käyttämään. Pääoma-arvoja laskettaessa käytetään suoraan sairastumisintensiteettejä, joten tältä osin on perusteltua tarkastella myös suoraan intensiteettejä. Myös tiheysfunktioita tullaan käyttämään.

Mallin selittäviä muuttujia valittaessa käytetään ns. Lagrange multiplier -testiä (ks. Rao [1973]). Cox, Oakes [1980, s.35] on selostanut muita mahdollisia testejä. Testi pohjautuu ML-estimaattoreiden asymptoottisiin ominaisuuksiin.

Mallin yleiseen sopimiseen soveltuvia testejä ovat mm. Kolmogorev-Smirnov -testi sekä Pearsonin yhteensopivuustesti (ks. esim. Lindgren [1976]). Molempien testien taustajatuksena on laskea erotus havaitulle ja sovitetulle kertymäfunktioille, joten ne soveltuvat erinomaisesti vaikkakin ovat hankalia suurilla otoskoilla. Tietysti myös likelihood ratio testi soveltuu mallin yleisen sopimisen testaamiseen.

Mallin diagnostiikka on aina tärkeää. Residuaalitarkastelut tullaan tekemään ns. deviance-residuaaleja sekä käyttämällä elossaolotodennäköisyyksiä. Sairastumisintensiteettejä tullaan myös tarkastelemaan residuaalien avulla. Elinaikamallien diagnostiikkaa sekä muita residuaalitarkasteluita on selostettu Klein, Moeschbergerissä [1997].

T. Tulokset

Tässä luvussa esitellään mallit, jotka on sovitettu käytettävissäoleviin aineistoihin. Tässä luvussa esiteltävien mallien sovittaminen on tehty SAS 6.0:n phreg ja lifereg ja SURVO 84C histo-ohjelmilla sekä EXCEL 5.0 taulukkolaskimen solver-ohjelmalle tehdyillä ohjelmilla.

T.1. Kuolevuus ennen sairastumista

Kuten jo aiemmin todettiin tähän mallinosaan ei sovitettu uutta aineistoa, vaan päätettiin käyttää kuolevuutena TEL:n perusvakuutuksessa jo olevaa kuolevuutta, jota on käytetty mm. vanhuuseläkeliiikkeessä. Erilaiset kuolevuuskäsitteet johtavat vain sekaannuksiin ja toisaalta - kuten myöskin aiemmin todettiin - aineistona voidaan käyttää vain samanlaista aineistoa, jota käytettäisiin vanhuuseläkeliiikkeen kuolevuuden tutkimiseksi.

Kuolevuuden tutkimiseksi poimittiin Eläke-Varman kannasta otos, johon tuli 60 042 henkilöä ja joista 438 oli kuollut. Poiminnassa pyrittiin etsimään ne tapaukset, jotka olisivat sairastuessaan saaneet juuri Eläke-Varmasta työkyvyttömyyseläkkeen ns. viimeisen laitoksen periaatteen mukaisesti. Poimitut tapaukset luokiteltiin syntymävuoden ja sukupuolen mukaan. Poimintaan otettiin mukaan kaikki 1.1.1962 -31.12.1996 aikana vakuutetut työsuhteet. Vanhimmat henkilöt olivat syntyneet 1897 ja nuorimmat 1983, joten luokkia tuli yli 170. Lisäksi aineistoa pienensi se, että tässä tehtävään tarkasteluun otettiin vain ne, jotka ovat korkeintaan 65- ja vähintään 15-vuotiaita tänä tarkasteluaikana. On selvää, että näin suppeasta aineistosta - 438 kuolemaa - ei riitä tapauksia jokaiseen luokkaan.

Kuolevuuden tarkastelemiseksi laskettiin luvun A.1 mukaiset epäparametriset kuolevuusestimaatit tapaa A käyttäen. Kuolevuutta ei suoraan verrattu vaan sitä käytettiin yhden vuoden elossaolotodennäköisyyden $\hat{q}_x$ laskemiseksi. TEL:n perusvakuutuksen yleisessä laskuperusteessa kuolevuus on määritelty

$$\mu(x) = (a1) \cdot e^{(a2)\{x+(b2)\}} \quad , \text{ jolloin } S(x | (b2)) = e^{-\frac{(a1) \cdot e^{(a2)(b2)} \cdot (e^{(a2) \cdot x} - 1)}{(a2)}} \quad \text{ja}$$

$$q(x) = \frac{1 - F(x)}{1 - F(x-1)} = e^{-\frac{(a1) \cdot e^{(a2)(b2)} \cdot e^{(a2) \cdot x} \cdot (1 - e^{-(a2)})}{(a2)}} .$$

Liitteen T kuvioissa T1.1 ja T1.2 on piirretty koko aineistolle mallin mukainen $q(x)$ (viiva) ja aineistosta laskettu $\hat{q}_x$ (pisteet) erikseen miehille ja naisille. Kuviossa kullekin syntymävuodelle on oma merkkinsä. Kuten tunnettua kuolevuus vaihtelee (b2) muuttujan eli syntymävuoden mukaan mistä johtuen kuvioissa on neljä erillistä viivaa. Liitteen T kuvioissa T1.3 ja T1.4. on piirretty muutamille ikäluokille mallin mukaiset elossaolotodennäköisyydet ja havainnoista saadut kaavan (A.3) mukaiset epäparametriset estimaatit. Tähän kuvioon on otettu mukaan vain muutama ikäluokka. Mallin mukaiset elossaolotodennäköisyydet on laskettu niin, että ikäluokan tullessa TEL:ssä vakuutetuksi elossaolotodennäköisyys on täysi yksi eli ehdolla, että ollaan selvitty tähän ikään - aivan kuten epäparametriset estimaatitkin on laskettu - vertailun mahdollistamiseksi.

Kuvioissa T1.1. ja T1.2 mallin yhden vuoden elossaolotodennäköisyys pysyy havaintojen keskellä. Tosin havainnoista ei voi päätellä ikäalueen loppupään taittumista, jonka malli tekee. Näin suppeasta aineistosta ei voi kuitenkaan päätellä vielä mitään. Kuvioden T1.3. ja T1.4. elossaolotodennäköisyydet näyttäisivät kaartavan mallin mukaisesti alaspäin. Mallin sopimista ei testattu formaalisti.

T.2. Sairastuvuus

Sairastuvuuden mallittamiseksi oli aineistona käytettävissä 117 vuosien 1994-1995 aikana Eläke-Varman kannasta löytynyttä sairastapausta. Mukaan voitiin ottaa vain kunkin kuun 13:ntena päivänä syntyneet, koska seurannasaoloaikaan, johon tapauksia verrataan, oli työsuhdekannasta poimittu vain nuo 13:ntena päivänä syntyneet. Seurannassaoloaikaa oli noina vuosina näillä henkilöillä yhteensä 15078 työvuotta. Sairastuvuus on siis alle prosentin luokkaa. Aineisto luokiteltiin iän tapahtumavuonna ja sukupuolen mukaan. Aineistosta laskettiin tavan B mukaiset epäparametriset estimaatit kaavojen (A.1) - (A.3) mukaan. Kävi kuitenkin selväksi, että luokkia täytyi yhdistää, mikä selviää myös liitteen T kuviossa T2.1, johon on piirretty ikäluokka- ja sukupuolikohtaiset havaitut intensiteetit. Kuten kuviosta näkyy vaihtelu ikäluokkakohtaisten intensiteettien välillä on melkoista. Luokkia täytyi yhdistää. Liitteen T kuviossa T2.2 on piirretty kaavojen (S.5) ja (S.6) innoittamat hajontakuviot. Ainakaan sukupuolella ei tuntuisi olevan juuri vaikutusta näiden mallien mukaiseen intensiteettiin. Yksittäiset ikäluokat koottiin viisivuotisikäluokkiin vaihtelun vähentämiseksi. Näin saadun uuden jakauman mukaiset intensiteetit on piirretty liitteen T kuvioon T2.3. Lisäksi laskettiin vastaavat elossaolotodennäköisyydet sekä tiheysfunktioestimaatit, joiden avulla laskettiin uusi frekvenssijakauma niille 117 sairastapaukselle nyt niin, että katotapausten paino tuli oikein huomioiduksi. Alkuperäinen ja uusi frekvenssijakauma on taulussa T2.1. Tätä uutta jakaumaa käytettiin, kun kaavan

(A.5) mukaisen maksimum likelihood estimoinnin avulla sovitettiin jakaumia tähän havaintoaineistoon. Tähän havaittuun jakaumaan yritettiin sovittaa kaikkia luvun S.1 mukaisia jakaumia eksponenttijakaumaa lukuunottamatta.

Taulu T2.1 Sairastumisiän frekvenssijakauma

ikäalue		alkuperäinen jakauma	korjattu jakauma
15 -	20	0	0
21 -	25	1	1
26 -	30	5	3
31 -	35	5	3
36 -	40	9	6
41 -	45	14	8
46 -	50	28	17
51 -	55	32	26
56 -	60	20	30
61 -	65	3	22

Sovitetuista jakaumista Weibull-jakauma putosi pois heti aluksi. Se ei konvergoinut SURVO 84C eikä EXCEL -ohjelmilla tehdyissä yrityksissä. Liitteen tauluun T2.2 on koottu tulokset log-logistisen sekä Gompertz-Makeham jakaumien ML-estimoinnin tulokset. Log-logistisen jakauma osalta saatiin ML-tulokset sekä SURVO 84C:llä että EXCELiin tehdyllä ohjelmalla. Gompertz-Makeham jakauman sovitteesta on sanottava, että se konvergoi aina niin, että parametri a läheni nollaa, joten tulokset ovat Gompertz-jakauman mukaiset.

Koska sovitteet eivät tunnu sopivan kovin hyvin - Pearsonin yhteensopivuustesti ei jää kummallakaan jakaumalla voimaan - tehtiin vielä kokeilu kahden normaalijakauman sekoitteella. Myös tämän kokeilun tulokset ovat taulussa T2.2; nyt yhteensopivuustesti jää voimaan vaikka parametrit on estimoitu tästä aineistosta. Lisäksi tauluun T2.2 on koottu kullekin jakaumalle tehdyt ns. ad hoc jakaumasovitteet (jotka siis on tehty muuten kuin ML-estimoinnilla). Liitteen tauluun T2.2 on laskettu myös Kolmogorov-Smirnov testisuure jakauman sopimiselle. Liitteen T kuvioissa T2.4 - T2.11 on piirretty näiden jakaumien sovitteita havaittujen epäparametrusten estimaattien kanssa.

Sovitetuista jakaumista tuntuu parhaiten sopivan kahden normaalijakauman sekoitus. Log-logistinen sekä Gompertz-Makeham jakaumien vaikeutena on tulla alaspäin heti ikäluokan 56 - 60 jälkeen - normaalijakaumat kykenevät tähän. Liitteen T kuvioon T2.12 on piirretty kahden normaalijakauman sekoitteen mukaisen intensiteetin ennuste ja havaitut intensiteetit nyt 1-vuotisikäluokittain sekä sukupuolet erikseen. Kuvioon T2.13 on piirretty

havaitun ja mallin mukaisen intensiteetin ero. Ainakin tämän kuvion perusteella tuntuisi siltä, että jotain jää vielä mallittamatta vaihtelun kasvaessa iän mukana. Nyt käytettävissä oleva aineisto ei kuitenkaan anna tähän mahdollisuutta.

T.3 Parantuvuus ja kuolevuus työkyvyttömyyseläkkeeltä

Eläke-Varman vuosilaskennoista 1994-1995 poimitusta aineistosta löytyi yhteensä 27 828 työkyvyttömyyseläkkeelläolijaa (ml. kuntoutustukitapaukset), joista tarkasteluajana 1.1.1994 - 31.12.1996 parani 1327 ja kuoli 540. Tarkasteluajana parantuneista 522 on määräaikaisen työkyvyttömyyseläkkeen (tai kuntoutustuen) päättymisiä, jotka luvun A.3.2 mukaan on määritelty parantuneiksi. Aineistossa on siis ns. katotapauksia yli 93 prosenttia. Varhaisimmat sairastumistapaukset ovat 1960-luvun alusta, joten pisin työkyvyttömyysaika on yli 35 vuotta. Seurannassaoloaika ei kuitenkaan voi ottaa huomioon kuin hetkestä 1.1.1994 lähtien, koska tätä hetkeä aiemmin alkaneilla tässä otoksessa olevilla eläkkeillä ei ole ollut vaaraa, että ne päättyisivät ennen hetkeä 1.1.1994. Vain tämän hetken jälkeen voimassa olleet eläkkeet ovat olleet vuosilaskennoissa vuosina 1994-1996 mukana. Vuosilaskennoissa on tosin mukana jokunen juuri ennen kyseistä aikaa päättyvä eläketapaus - ne ovat 'myöhästyneet' edellisestä vuosilaskennasta ja nyt ne on poistettu tästä tarkastelusta. Seurannassaoloaika poimintaan tuli yhteensä 72 586 työkyvyttömyyseläkevotta, joten keskimääräinen parantuvuus on alle prosentin ja kuolevuus lähes 2 prosenttia ja siis pisin seurannassaoloaika on 3 vuotta. Taaskin yksilölliset varhaiseläkkeet jätettiin tarkastelusta pois. Liitteen S kuvioihin T3.1 - T3.3 on piirretty SAS:n lifetest ohjelmalla saadut epäparametriset intensiteetit koko aineistosta.

Poiminnassa mukaan otettiin seuraavat tiedot: syntymäaika, sukupuoli, eläketapahtumapäivä, eläkkeen päättymispäivä ja -syy. Näistä muuttujista muodostettiin seuraavat liitteen T taulukon T3.1 mukaiset muunnokset ns. potentiaalisiksi selittäviksi muuttujiksi, tapahtuman indikaattorimuuttujiksi sekä seurannassaoloajaksi. Malli sovitettiin suoraan esityksen (A.6) mukaiseen likelihoodtarkasteluun. Sovittaminen tehtiin SAS 6.0 phreg ja lifereg ohjelmilla, joista edellinen laskee ns. suhteellisen vaaran malleja ja jälkimmäinen kiihtyneen tapahtuma-ajan malleja. Tapahtuman indikaattorimuuttujia tehtiin kolme: yksi kuolemislle, yksi parantumislle ja yksi näiden yhdistelmälle. Indikaattorimuuttuja siis kertoo onko kyseessä oleva tapaus tapahtuma eli esimerkiksi kuolema vai katotapahtuma eli esimerkiksi vanhuuseläkkeelle siirtyminen tai eläkkeen jatkuminen yli hetken 31.12.1996. Kiihtyneen tapahtuma-ajan mallia vaatii esityksen (S.9) mukaan virhejakautuksen, joita lifereg ohjelmassa oli valittavana useita.

Soveltuvimmat jakaumat, joita kokeiltiin, olivat Weibull-, log-logistinen sekä lognormaalijakauma.

Jakauman valitsemiseksi sovitettiin aluksi seurannassaoloajalle puhdas jakauma-malli, jossa ei ole ns. selittäviä tekijöitä mukana. Liitteen taulukkoon T3.2 on kerätty kullakin tapahtumaindikaattorimuuttujalla saadut tulokset. Näistä lognormaalijakauma tuntuu sopivan heikoiten ja Weibull- ja log-logistisen jakauman tulokset näyttävät kovin samanlaisilta - erot ovat kuitenkin pieniä.

Varsinaisten selittävien tekijöiden liittäminen malliin tehtiin SAS 6.0 phreg-ohjelmalla, joka laskee ns. suhteellisen vaaran malleja. Saatuja tuloksia ei kuitenkaan suoraan voi käyttää hyväksi pääoma-arvoja laskettaessa. Phreg-ohjelma valittiin siksi, että se käyttää ns. perusjoukon intensiteettinä $\mu_0(x)$ aineiston havaittuja intensiteettejä. Jos lifereg-ohjelma antaa tuloksena erityisesti Weibull-jakaumalla samanlaisia parametriarvoja, voidaan olla ainakin hieman vakuuttuneempia siitä, että valittu perusjakauma on juuri oikean suuntainen eli havaitun jakauman mukainen. Lifereg-ohjelman parametriesitys täytyy muuntaa luvun S.3 mukaan vastaamaan suhteellisen vaaran mallin mukaista parametriesitystä.

Sairastumisintensiteettiä mallitettaessa valituksi tuli malli, jossa selittävinä tekijöinä ovat ikä tapahtumavuonna (TVJ), syntymävuosi (SVJ) sekä indikaattorimuuttujat syntymävuosi 1940-, 1950 ja 1960-luvulla (SVL1940, SVL1950, SVL1960). Liitteen T taulukkoon T3.3 on koottu Weibull-, log-logistisen sekä suhteellisen vaaran mallin mukaiset parametriesitykset. Kuten tuloksista huomataan, ovat suhteellisen vaaran mallin sekä Weibull-jakauman mukaisen kiihtyneen tapahtuma-ajan mallin mukaiset parametrit käytännössä samanlaiset sekä Weibull-jakauman ja log-logistisen jakauman parametrit samat. Tämä vahvistaa uskoa siihen, että valittuja jakaumia voitaneen käyttää. Tulosten mukaan näyttäisi siltä, että 1940-, 1950- ja 1960-luvulla syntyneillä on merkittävästi suurempi 'riski' parantua kuin muilla eli lähinnä 1930- ja 1970-luvulla syntyneillä. Yksi ikävuosi ikään tapahtumavuonna lisää hieman riskiä parantua ja syntymävuoden kasvaminen 'paljon'. Liitteessä S on SAS-ohjelmien tulosteet näistä malleista.

Kuolemisintensiteettiä mallitettaessa valituksi tuli malli, jossa selittävinä tekijöinä ovat ikä tapahtumavuonna (TVJ), syntymävuosi (SVJ) ja indikaattorimuuttujat sukupuoli (SPN) ja ns ei lähellä eläkeikää tapahtuvan eläketapahtuma-ajan muuttuja (TVL4) sekä kovarianssimuuttujat TVJ*SPN ja SVJ*SPN. Kovarianssimuuttujat kertovat kuinka paljon naisten vaara kuolla lisääntyy iän tapahtumavuonna ja syntymävuoden lisääntyessä miesten vastaavan vaaran lisäksi. Sukupuolimuuttuja kertoo paljonko on miesten ja naisten välinen tasoero vaaralle. Taas liitteessä T taulussa T3.3 on parametrit kolmesta sovitetusta mallista.

Tulosten mukaan näyttäisi siltä, että naisten riski kuolla olisi pienempi kuin miesten, samoin lähellä eläkeikää sairastuminen on vaarallisempaa kuin aiemmin. Iän tapahtumavuonna ja syntymävuoden kasvaessa riski pienenee, kuitenkin naisten osalta riski kasvaa hieman nopeammin kuin miesten. SAS tulosteet on liitteessä S.

Yllä sovitettuja malleja voitaisiin soveltaa esimerkiksi (X.9) mukaista odotusarvoa laskettaessa. Koska mallissa olisi kaksi intensiteettiä ja useita parametreja, ei odotusarvo olisi todennäköisesti integroitavissa suljetussa muodossa. Osittain tästä käytännöllisestä syystä päätettiin sovittaa malli myös yhdistetylle tapahtumalle, jossa tapahtumana on työkyvyttömyyseläkkeen ennaikainen päättyminen joko paranemiseen tai kuolemaan. Tämä tehtiin siitäkin huolimatta, että edellä esitellyt tulokset vahvasti viittavat siihen, että kyseessä olevat tapahtumat noudattavat erilaista kulkua - todennäköisyysjakaumaa.

Yhdistettyä intensiteettiä mallitettaessa valituksi tuli malli, jossa selittävinä tekijöinä ovat ikä tapahtumavuonna (TVJ), syntymävuosi (SVJ) ja indikaattorimuuttujat sukupuoli (SPN), syntymävuosi 1930- ja 1940-luvulla sekä kovarianssimuuttujat TVJ*SPN ja SVJ*SPN. Naisten vaara eläkkeen päättymiselle on taas pienempi kuin miesten, samoin 1930- ja 1940-luvulla syntyneillä muihin verrattuna. Iän tapahtumavuonna ja syntymävuoden kasvaessa vaara lisääntyy ja naisilla nopeammin kuin miehillä. SAS-tulostus on liitteessä S.

Sovitetun mallin vertailemiseksi nykyiseen perusteeseen on liitteen T kuvioon T3.1 piirretty log-logistisen mallin ennuste kaavan (X.4) mukaisesta odotusarvosta ja nykyisen laskuperusteen vastaava pääoma-arvo heti alkavalle eläkkeelle ($e=0$, korkoutuvuus = 0). Jotta kuviossa olisi myös havaintoja, on vuosilaskennoista poimitusta aineistosta poimittu taas 13:ntena päivänä syntyneet ja laskettu heidän tavan B mukaiset havaitut yhden vuoden eläkkeen päättävyyseriskit $\hat{q}_x$. Näitä yhteenlaskemalla saadaan estimaatti 'havaitusta odotusarvosta'. Kuvioon on piirretty ennusteen aikaa vain iästä 58 ikään 65. Tätä pidemmälle ajalle (ainakaan) ei mallia kannata käyttää, koska mallia sovitettaessa on kolme vuotta ollut pisin seurannassaoloaika. Näin ollen kuviossa on jo nyt havaintojen mukaan mallitettua aikaa vain kolme vuotta ja tätä pidemmät ajat eivät perustu havaintoihin vaan ovat ekstrapoloituja. Käytettävissä oleva aineisto ei mahdollistanut käyttökelpoisemman mallin sovittamista.

Yllä kuvatuista malleista on liitteen T kuvioihin T3.2 - T3.9 piirretty deviance residual kuviot, jossa ympyrällä on merkitty tapahtuma ja pisteellä katotapahtumat. Suhteellisen vaaran mallin mukaiset residuaalikuviot ovat käytännössä samanlaiset kuin kiihtyneen

tapahtuma-ajan mallissa. Kuvioissa silmiinpistävä piirre on se, että tapahtumat ja katotapahtumat erottuvat omiksi joukoikseen. Tämä lienee deviance residuaalien ominaisuus. Kuvioon T3.10 on piirretty eri elossaolotodennäköisyyksien arvolla deviance residuaalien arvoja. Tosin residuaalikuvioista on lisäksi sanottava, että parantumista mallitettaessa vielä parempi malli olisi ilmeisesti löydettävissä - tiettyä lineaarisuutta on havaittavissa. Muut mallit näyttävät olevan kunnossa.

T.4 Johtopäätöksiä

Valitettavasti on lopuksi todettava, että nyt sovitettuja malleja ei suositella suoraan käyttöönotettaviksi. Aineisto ei ollut tarpeeksi laaja kunnollisen sovellettavissa olevan mallin luomiseksi TEL:n perusvakuutuksen pääoma-arvojen laskemiseksi.

Kuolevuudelle ennen sairastumista ei pyrittykään tekemään uutta mallia vaan yritettiin vain tarkistaa onko nykyisen vanhuuseläkeliikeen laskuperuste kunnossa. Käytössä olevan aineiston perusteella ei näytä siltä, että laskuperuste ei ole kohdallaan. Sopimista ei kuitenkaan testattu.

Sairastumisintensiteetti mallitettiin suoraan todennäköisyysjakaumaan sovittamalla. Koska käytössä oli vain 117 vuosina 1994 -1996 sairastunutta, jouduttiin aineisto yhdistelemään hyvin suppeaksi. Havaitut intensiteetit saatiin lopulta tapaa B käyttäen eli yhdistelemällä eri vuosina saman iän täyttäneet, mutta vain viisivuotiskäryryhmittäin. Lopulliseksi malliksi löydettiin kahden normaalijakauman sekoite, joka näytti sopivan aineistoon myös tehtyjen testien mukaan. Kuitenkin residuaalikuvio oli huolestuttava. Aineiston vähyyden takia parempaa mallia ei voitu sovittaa.

Työkyvyttömyyseläkkeen päättymisen yritettiin mallittaa kattavimmin. Suurimmaksi ongelmaksi muodostui taas kuitenkin aineiston puutteellisuus. Käytössä oli vain kolme seurannassaolovuotta, kun työkyvyttömyysmallin pitäisi pystyä ennustamaan ilmiötä, joka saattaa jatkua vuosikymmeniä. Käytössä olevia pääoma-arvoja ei voi perustaa käytännössä pelkälle ekstrapoloinnille.

Käytössä olevaan aineistoon pystytään kuitenkin sovittamaan malli, joka voi toimia esimerkkinä, miten kyseessä olevaa ilmiötä voidaan mallittaa. Valituksi tuli kiihtyneen tapahtuma-ajan malli, jossa perusjakaumana on käytetty log-logistista jakaumaa ja jossa on liitettyinä selittäviä tekijöitä. Malli tuntuu sopivan käytössä olevaan aineistoon.

Tässä esitellyt mallit soveltuvat esimerkiksi siitä miten tätä ilmiötä voidaan eräällä tavalla

mallittaa. Myös muita mallittamistapoja ja -kriteereitä on löydettävissä. Nyt esiteltyjä malleja ei ole myöskään viimeistelty aivan loppuun juuri aineiston puutteellisuuden takia. Diagnostisia tarkasteluja voisi tehdä lisää: muun muassa mallin vaikutusvaltaisten havaintojen etsiminen olisi tärkeää. Lisäksi aineisto ei mahdollistanut luvussa A.1 esiteltyjen havaittujen intensiteettien tapojen A ja B vertailua, mikä olisi ollut tärkeää ja esimerkiksi jakaumasovitus olisi myös ollut kiintoisaa tehdä myös tälle työkyvyttömyyseläkkeelläoloajalle - ainakin vertailun vuoksi. Edelleen käyttöönotettavalta mallilta vaaditaan paitsi aineistoon sopimista, myös helppokäyttöisyyttä. Lopullista mallia valittaessa myös nämä käytännölliset seikat on huomioitava, joskus jopa aineistoon sopimisen hinnalla.

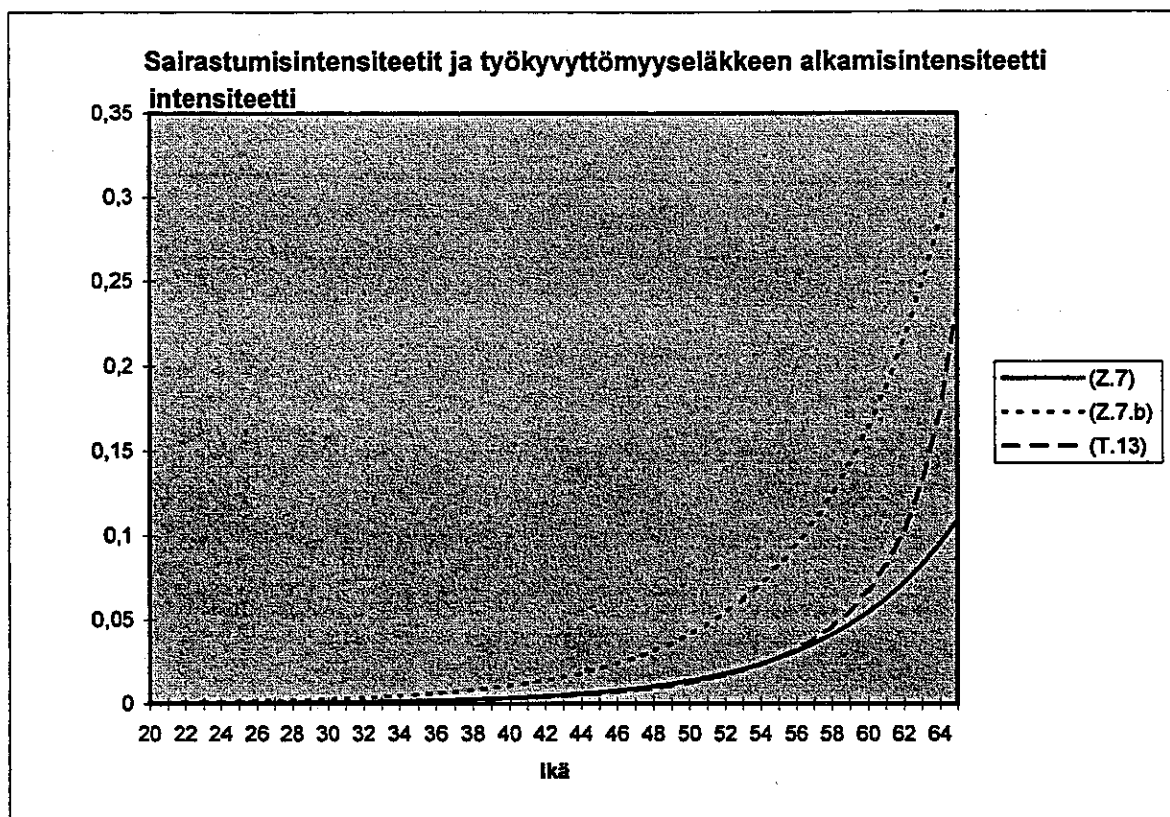
Lähdeluetteloon on merkitty kaikki lähteenä käytetyt teokset, joista käytettyjen mallien kannalta tärkeimmät ovat Cox, Oakes [1984] ja Kalbflesh, Prentice [1980].

Lähteet:

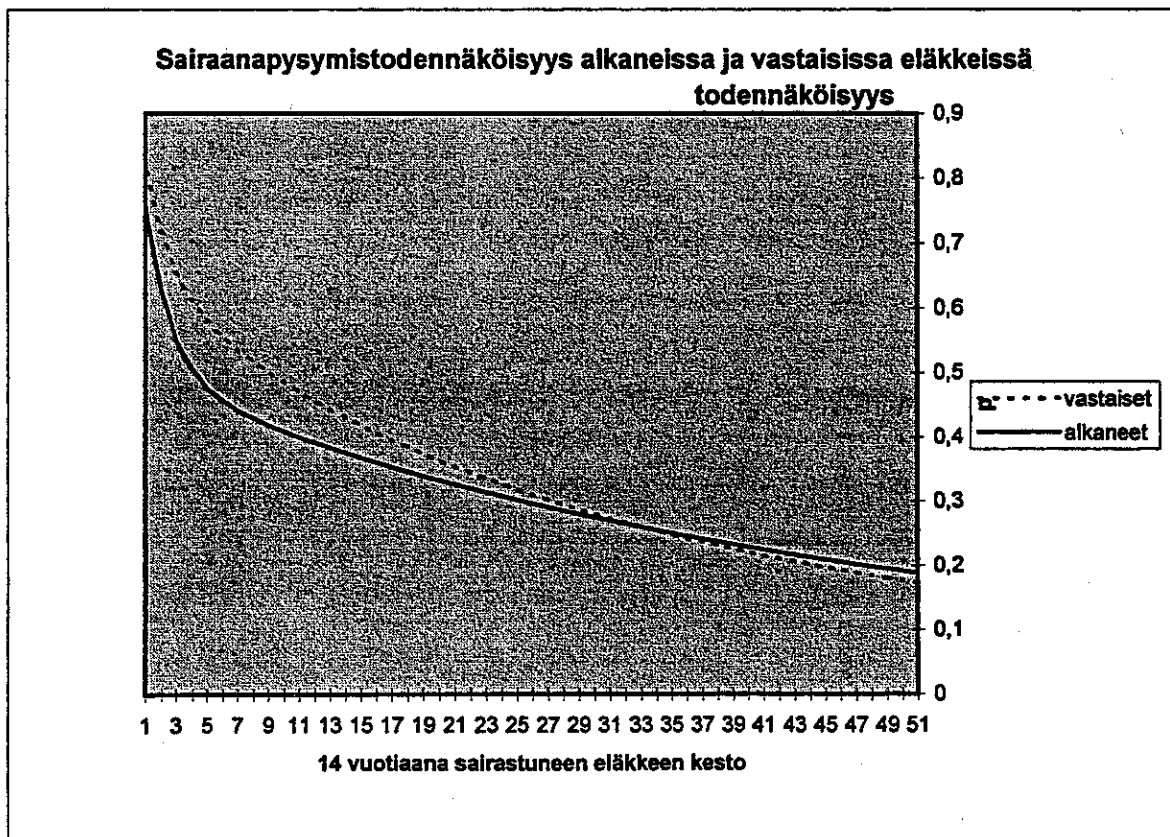
- Aranda-Ordaz (1983), An Extension of the Proportional-Hazards Model for Grouped Data, *Biometrics* 39,
- Bartholomew (1982), Stochastic Models for Social Processes,
- Bowers, Gerber, Hickman, Jones, Nesbit (1986), Actuarial Mathematics,
- Breslow (1972), Discussion on Professor Cox's Paper (ks. Cox [1972]), *Journal of the Royal Statistical Society, Series B* 34,
- Cox (1972), Regression Models and Life-Tables, *Journal of the Royal Statistical Society, Series B* 34,
- Cox, Oakes (1984), Analysis of Survival Data,
- Eland-Johanson, Johanson (1980), Survival Models and Data Analysis,
- Gauffin (1981), Työkyvyttömyysmalleista, SHV-harjoitustyö,
- Gauffin, Leinonen (1982), Työkyvyttömyysmalleista kirjassa Henki ja eläkevakuutuksen vakuutustekniikkaa,
- Gerber (1990), Life Insurance Mathematics,
- Greenwood, (1926), The errors of sampling the survivorship tables, Reports on Public Health and Statistical Subjects,
- Gregorius (1993), Disability insurance in The Netherlands, Insurance mathematics and economics 13,
- Hakulinen, Tenkanen (1986), Suhteellisen elossaololukujen regressioanalyysistä, Suomen Tilastoseuran vuosikirja 1986,
- Hakulinen (1983), Tapahtuman toteutumisen tai välttämisen yleisyyden arvioimisesta, Sosiaalilääketieteellinen Aikauslehti 1983:20,
- Hertzman (1993), Summary of the talk on 'Work of the CMI PHI Sub-Committee', Insurance mathematics and economics 13,
- Hytti (1993), Työkyvyttömyyseläkkeelle siirtymisen yhteiskunnalliset taustatekijät,
- Hännikäinen (1988), Työkyvyttömyysriski ja sen taustatekijät,
- Jasiulewicz (1997), Application of mixture models to approximation of age-to-death distribution, Insurance mathematics and economics 19,
- Kalbfleisch, Prentice (1980), The Statistical Analysis of Failure Time Data,
- Kaplan-Meier (1958), Nonparametric estimation from incomplete observation, J. Am. Statist. Assoc. 26,
- Klein, Moeschberger (1997), Survival Analysis, Techniques for Censored and Truncated Data,
- Laird, Olivier (1981), Covariance analysis of censored survival data using log-linear analysis techniques, J. Am. Statist. Ass 1981:76,
- Lassila (1977), TEL-työkyvyttömyystutkimus, SHV-harjoitustyö,
- Lindgren (1976), Statistical Theory,
- Manton, Stallard, Vaupel (1981), Methods for comparing the mortality experience of heterogeneous populations, Demography 1981:18,
- McCullagh, Nelder (1989), Generalized Linear Models,
- Microsoft (1993), EXCEL 5.0, User's Guide,

- Mustonen (1992), SURVO An Integrated Environment for Statistical Computing and Related Areas,
- Puranen, Data analyysi, HY Tilastotieteen laitoksen oppimateriaalia,
- Pitacco (1994), Actuarial models for pricing disability benefits: Towards a unifying approach, Insurance mathematics and economics 16,
- Portnoy (1997), Regression-quantile graduation of Australian life tables, 1946-1992, Insurance mathematics and economics 21,
- Puzey (1995), A reappraisal of the principle underlying the conventional actuarial estimator of q_x , Insurance mathematics and economics 17,
- Rao (1973), Linear Statistical Inference and its Applications,
- Renshaw, Haberman (1995), On the graduations associated with a multiple state model for permanent health insurance, Insurance mathematics and economics 17,
- SAS Institute Inc (1990), SAS/STAT User's Guide, 6,
- SAS Institute Inc (1996), SAS/STAT Software, Changes and Enhancements, 6.11,
- Sansom, Waters (1988), Permanent Health Insurance in the UK: The Mathematical Model and the Statistical Analysis of the Data, ICA 1988 Helsinki,
- Seber (1977), Linear Regression Analysis,
- Segerer (1993), The actuarial treatment of the disability risk in Germany, Austria and Switzerland, Insurance mathematics and economics 13,
- Scott (1988), On the Theory of Graduation of Mortality Tables, ICA 1988 Helsinki,
- TEL:n mukaisen vakuutuksen yleiset laskuperusteet (1990),
- TEL:n mukaisen perusvakuutuksen erityisperusteet (1997),
- Tibshirani, Ciampi (1983) A Family of Proportional- and Additive-Hazards Models for Survival Data, Biometrics 39,
- Tuomikoski (1986) TEL-työkyvyttömyysmalli: Z-pinnan parametrien määrittäminen pinnansovitusnäytteenä,
- Wolthuis (1994), Life Insurance Mathematics

Liite Z
Kuvio Z3.1

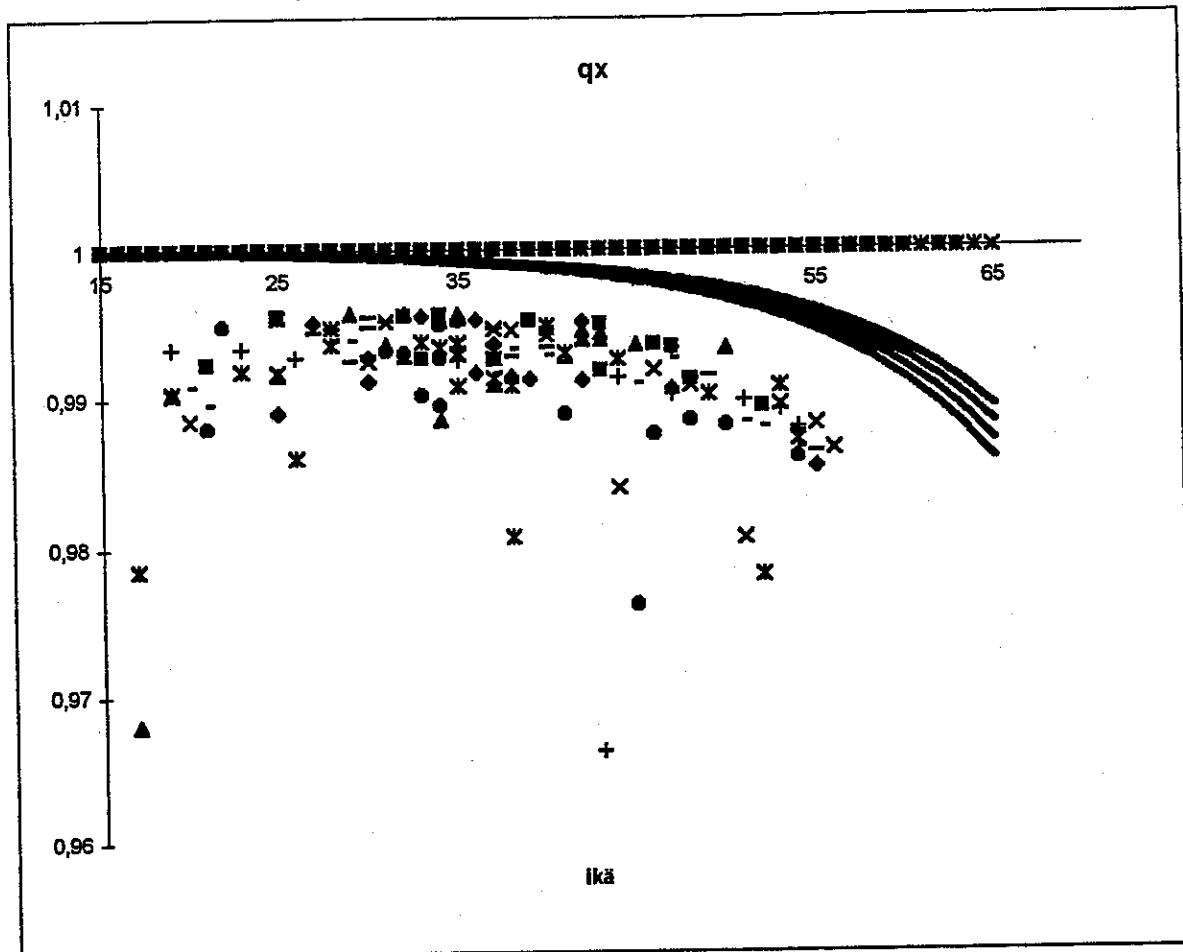


Kuvio Z3.2

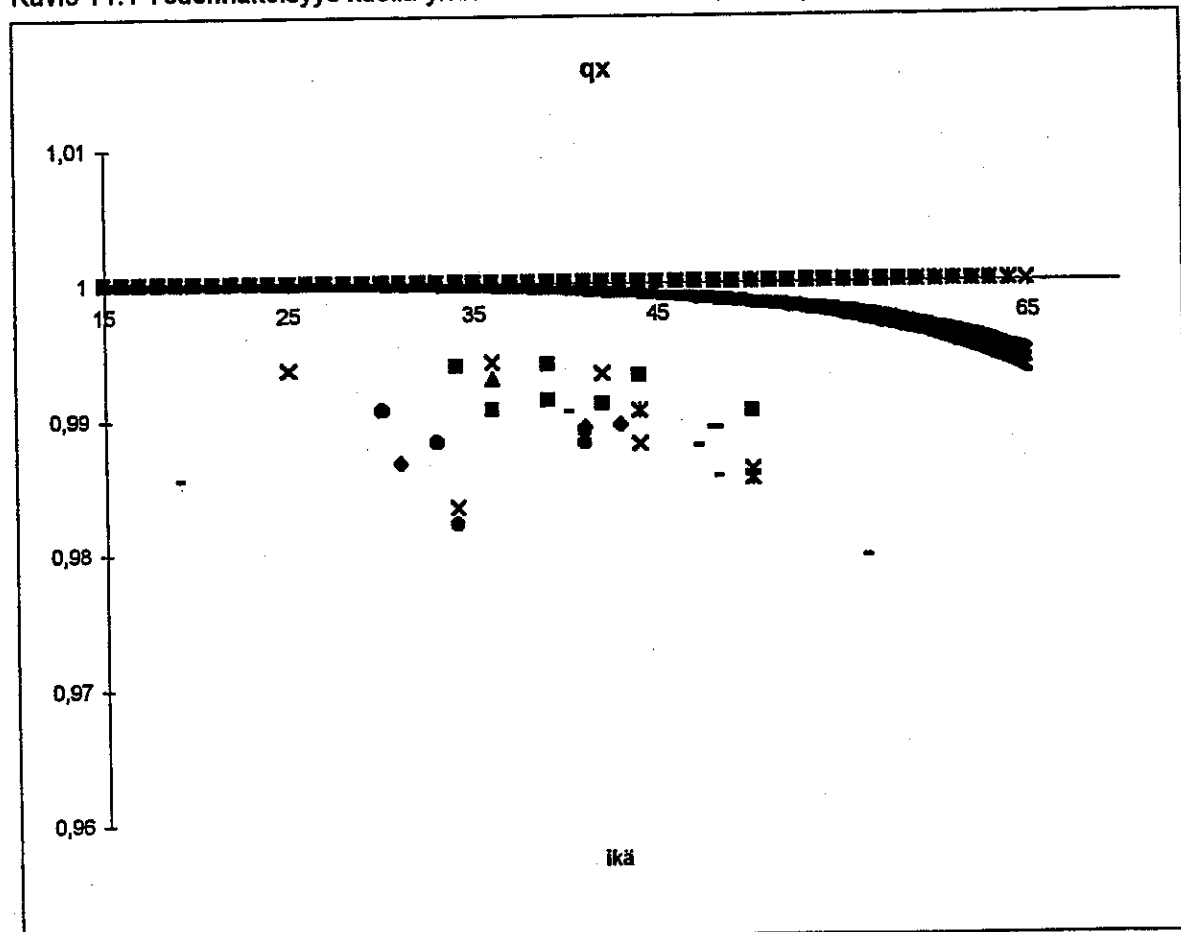


Liite T

Kuvio T1.1 Todennäköisyys kuolla yhden ikävuoden aikana, malli (viiva) vs havainnot (merkki), miehet



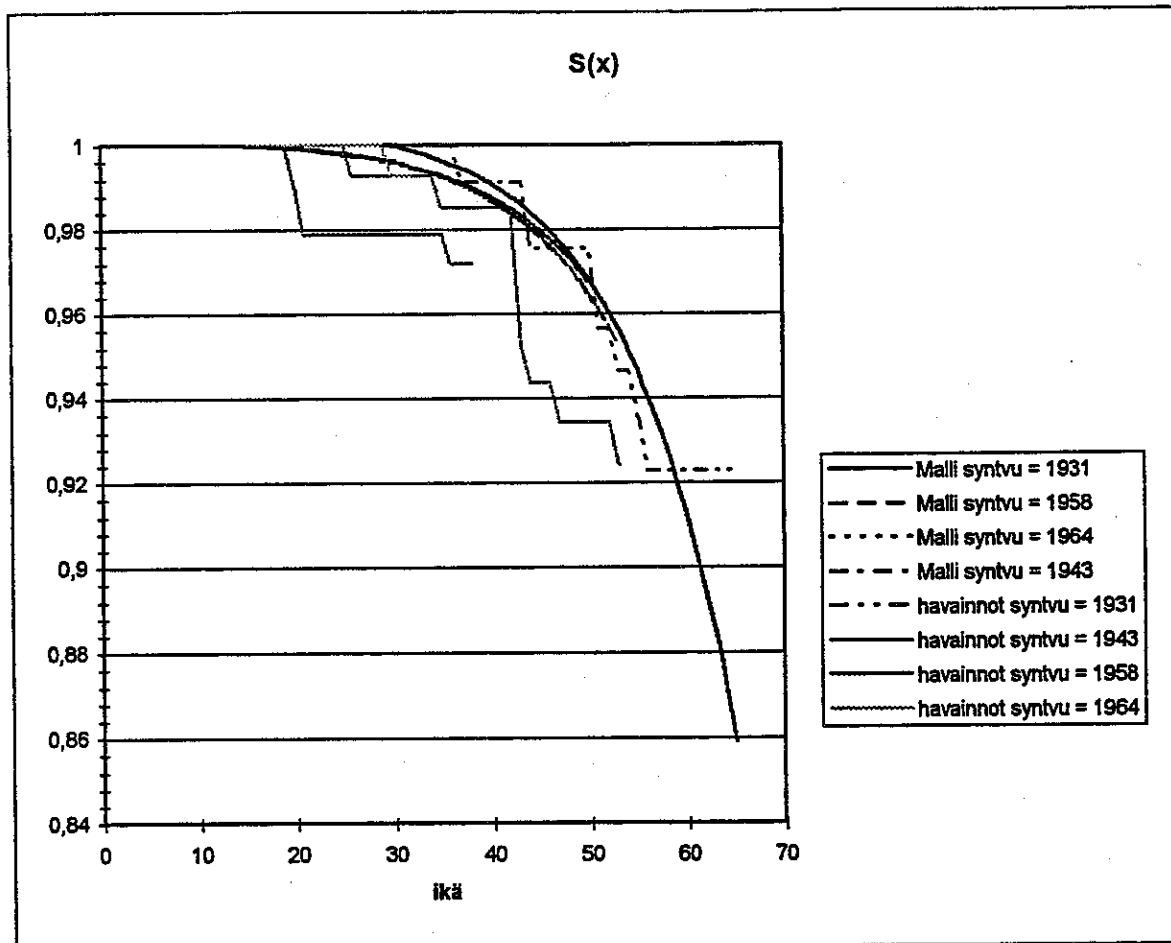
Kuvio T1.1 Todennäköisyys kuolla yhden ikävuoden aikana, malli (viiva) vs havainnot (merkki), naiset



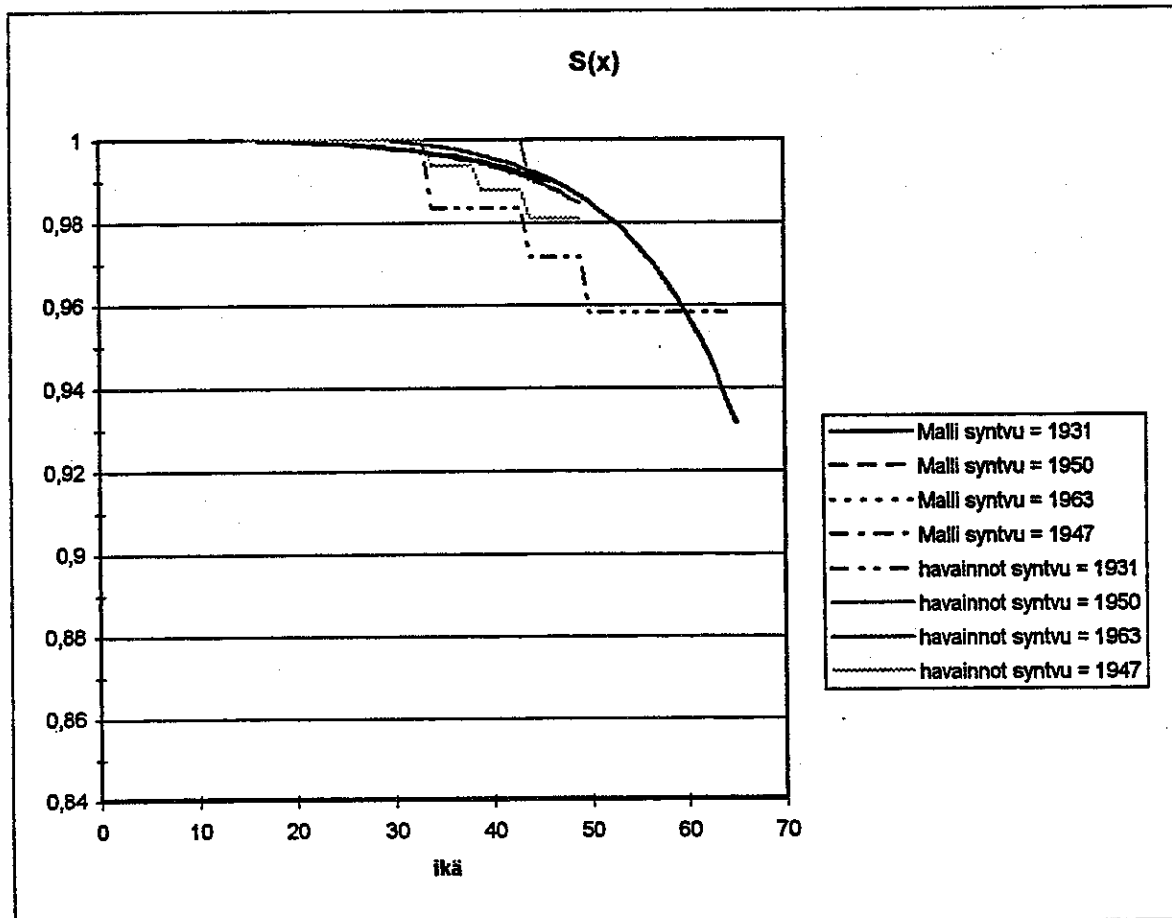
Havaituissa arvoissa kullakin syntymävuosilla on oma merkinsä.

Liite T

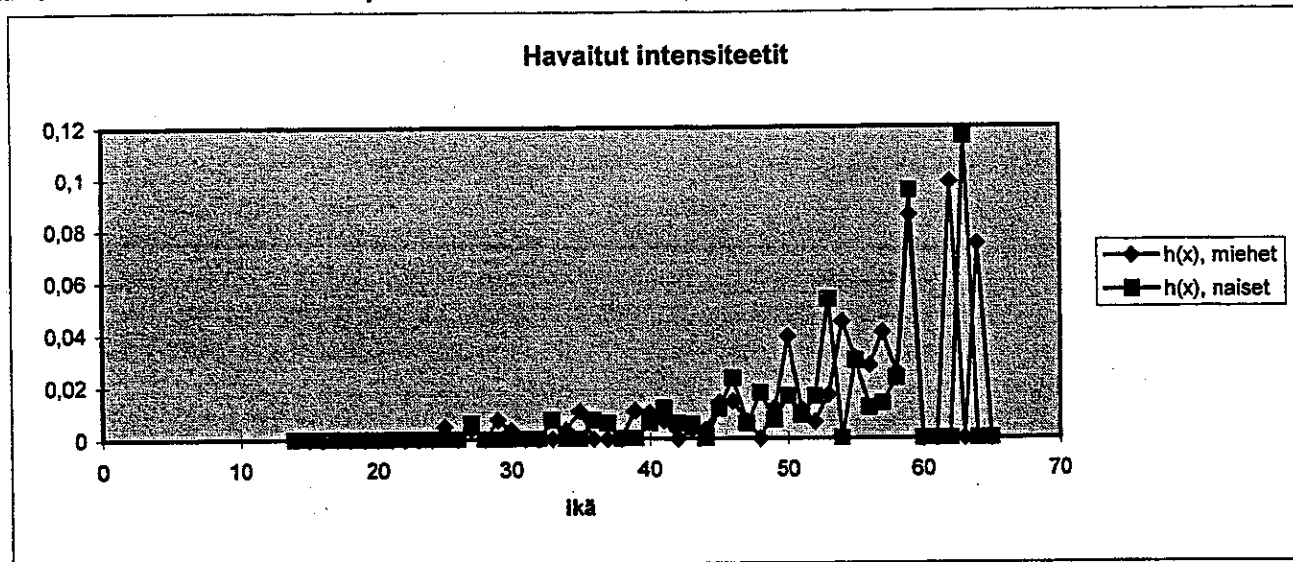
Kuvio T1.2 Elossaolotodennäköisyys tietyissä syntymävuosiluokissa, malli vs havainnot, miehet



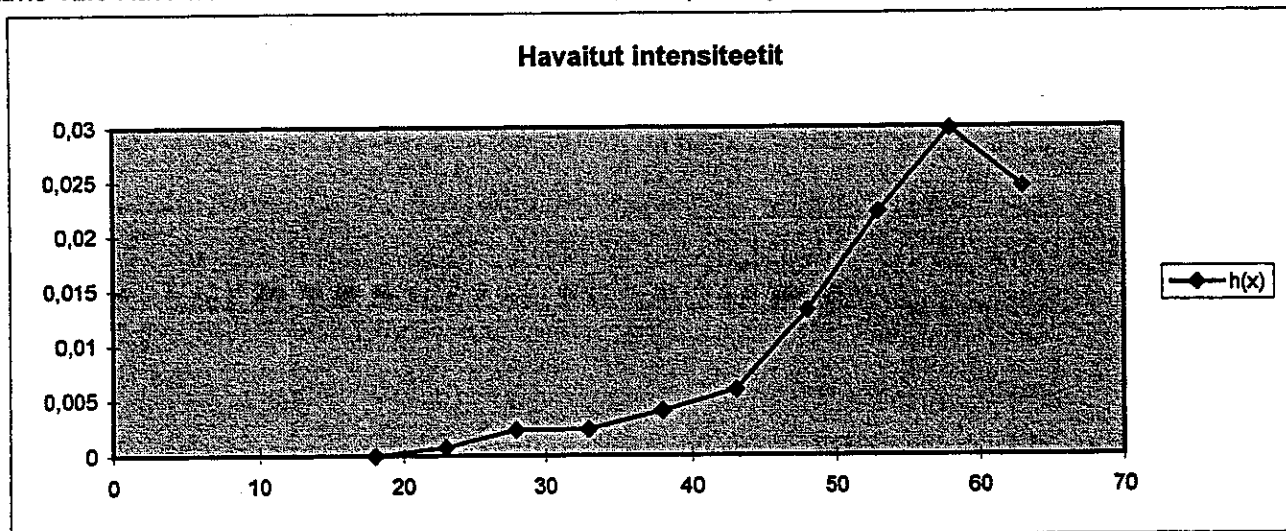
Kuvio T1.2 Elossaolotodennäköisyys tietyissä syntymävuosiluokissa, malli (P) vs havainnot, naiset



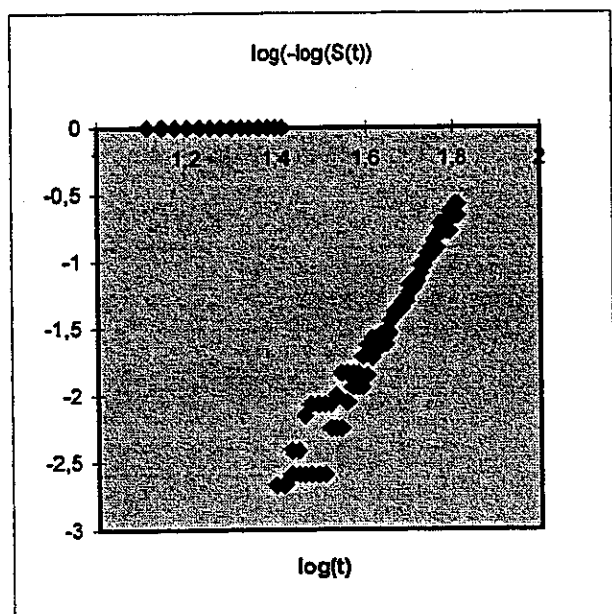
Kuvio T2.1 Havaitut intensiteetit yksivuotisikäluokittain, sukupuolet erikseen



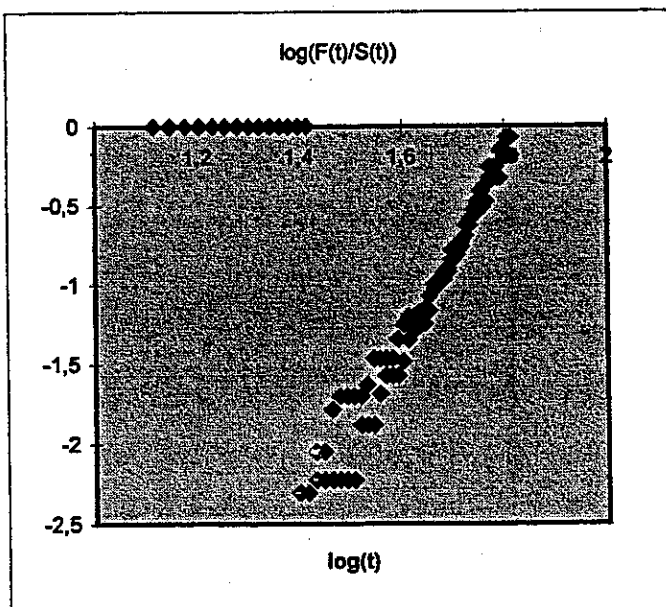
Kuvio T2.3 Havaitut intensiteetit viisivuotisikäluokittain, sukupuolet yhdessä



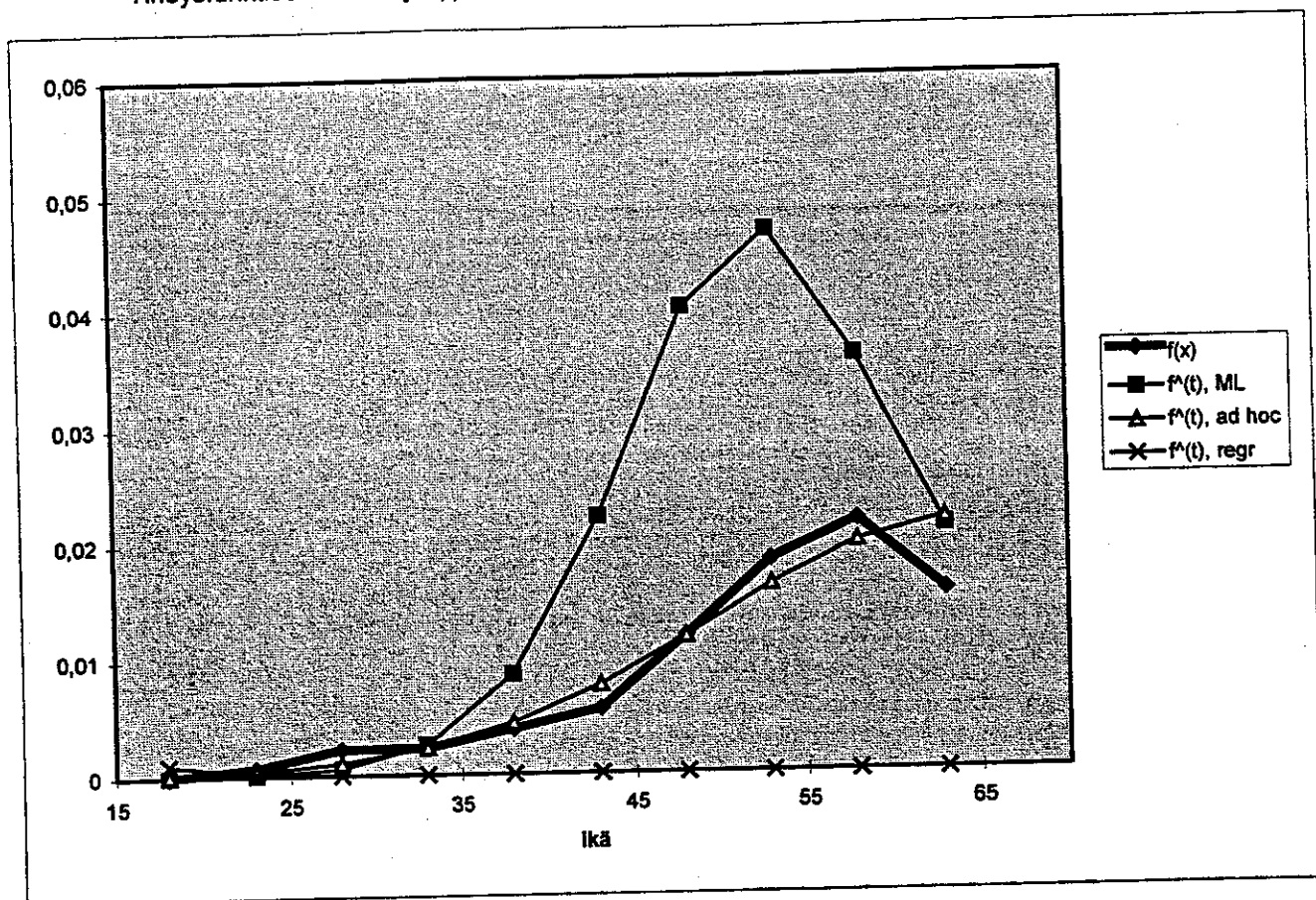
Kuvio T2.2 Havaitut $\log(-\log(S(x)))$ vastaan $\log(x)$



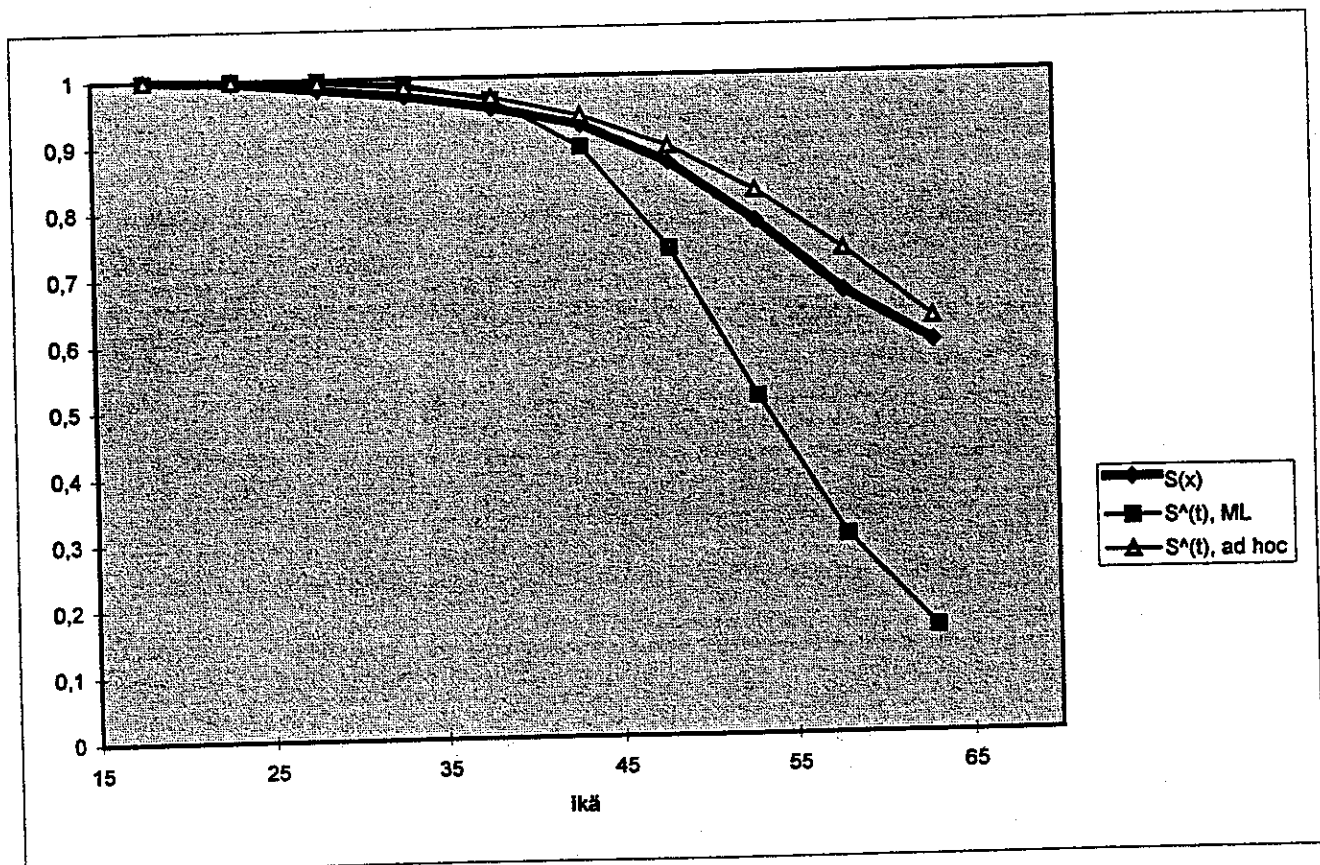
Havaitut $\log(F(x)/S(x))$ vastaan $\log(x)$



Huom. $y = 0$ arvot ovat i'istä joissa ei ole vielä sairastumisia.

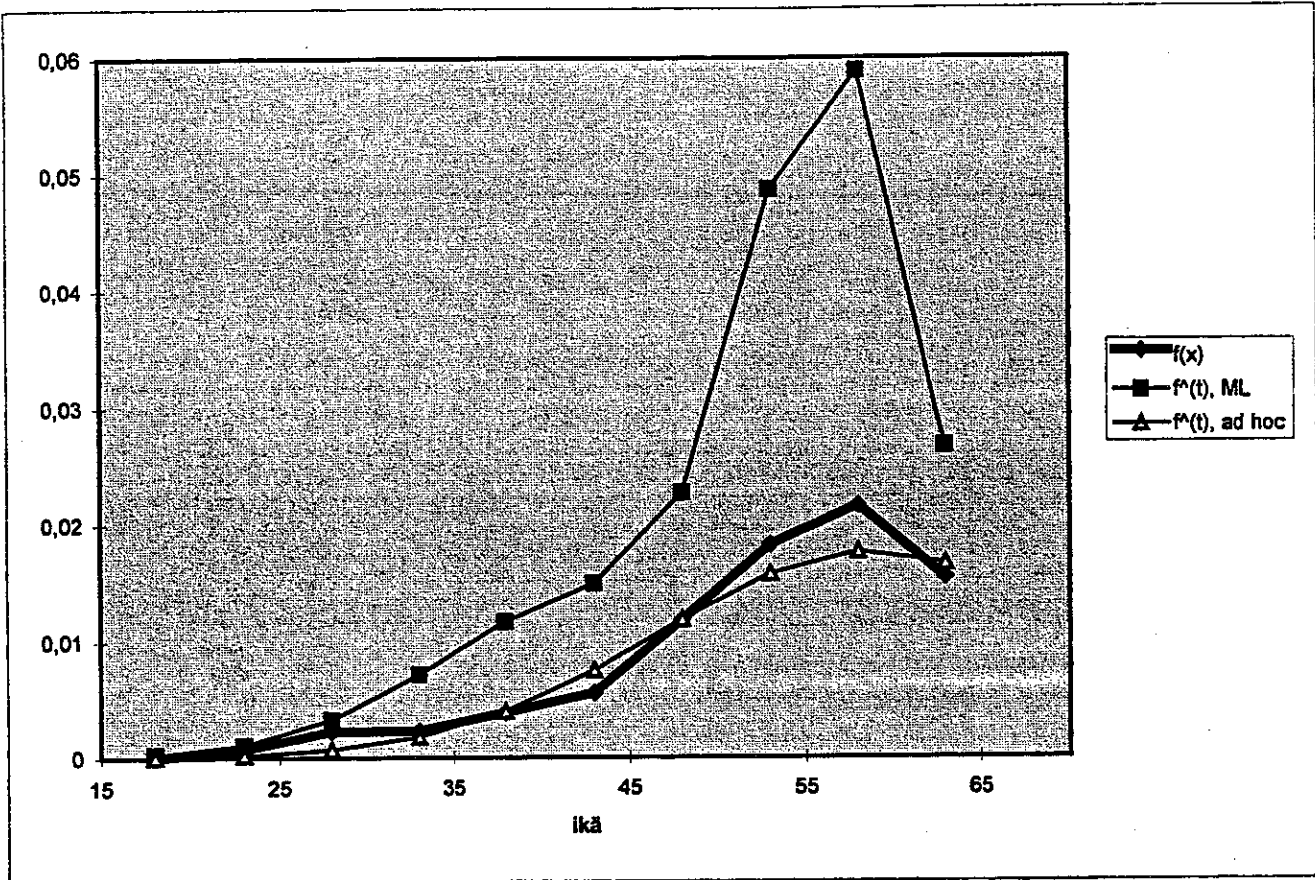


Kuvio T2.5 Eloisaolotodennäköisyysestimaatit [$S^{\wedge}(t)$ ML, $S^{\wedge}(t)$ ad hoc] verrattuna havaittuun $S(x)$

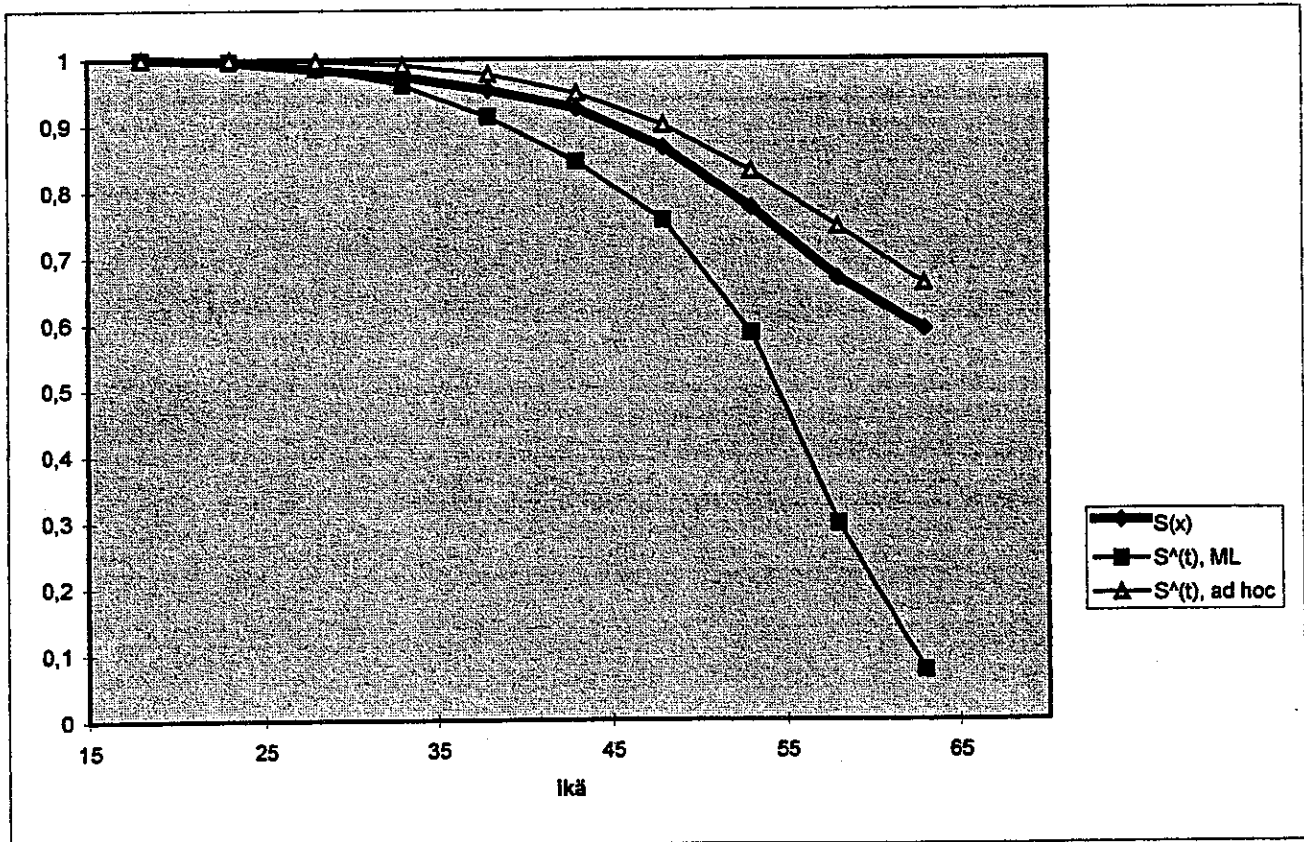


Liite T

Kuvio T2.6 Kahden normaalijakauman sovittaminen sairastumisiin jakaumaan
Tiheysfunktioestimaatit [$\hat{f}(t)$ ML ja $\hat{f}(t)$ ad hoc] verrattuna havaittuun $f(x)$

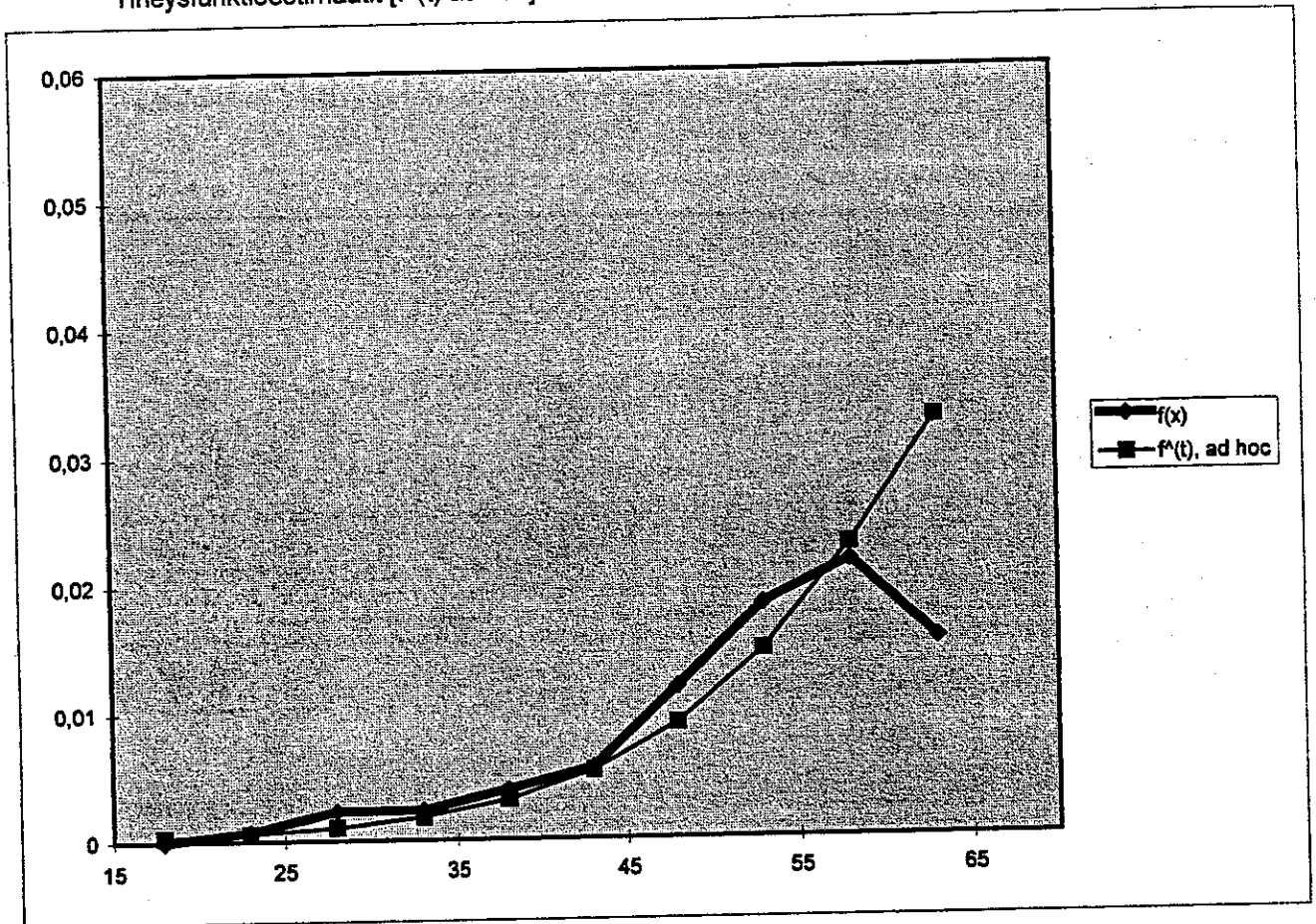


Kuvio T2.7 Eloissaolotodennäköisyysestimaatit [$\hat{S}(t)$ ML, $\hat{S}(t)$ ad hoc] verrattuna havaittuun $S(x)$

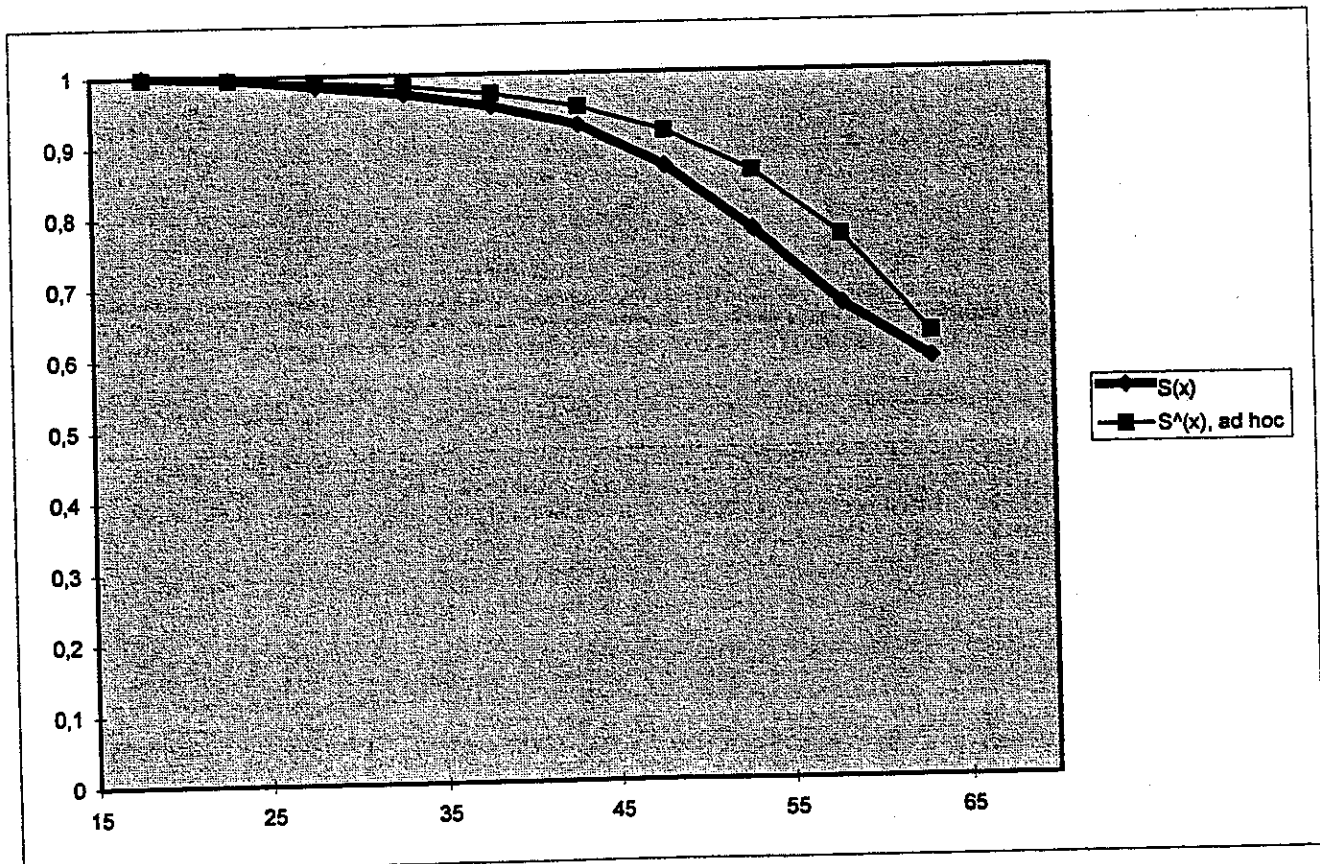


Liite T

Kuvio T2.8 Gompertz-jakauman sovittaminen sairastumisiin jakaumaan
Tiheysfunktioestimaatit [$f^{\wedge}(t)$ ad hoc] verrattuna havaittuun $f(x)$



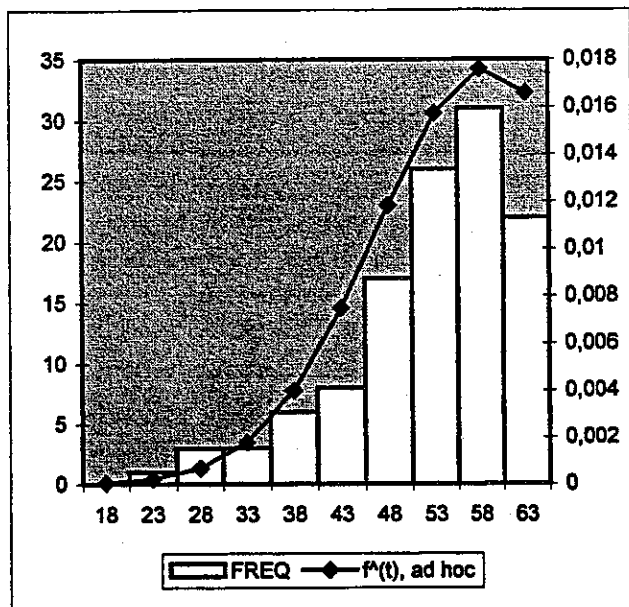
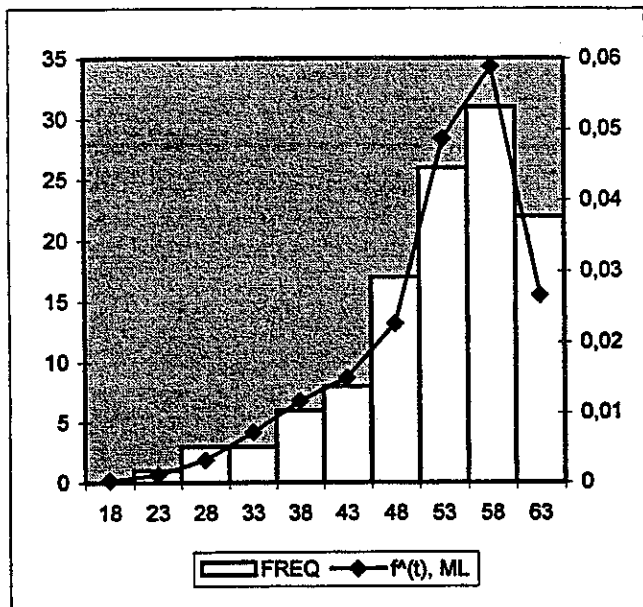
Kuvio T2.9 Eloisaolotodennäköisyysestimaatit [$S^{\wedge}(t)$ ad hoc] verrattuna havaittuun $S(x)$



Liite T

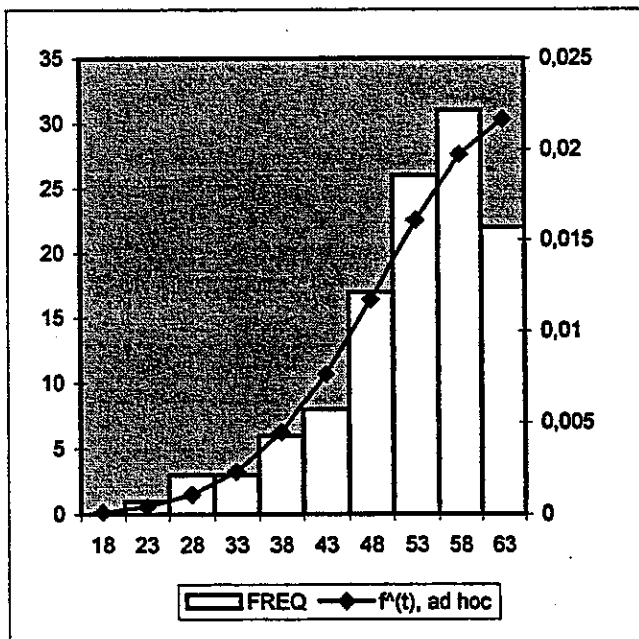
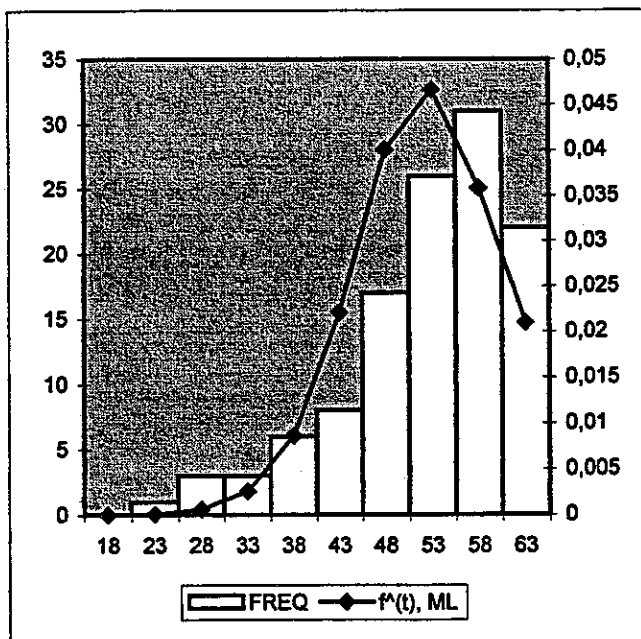
Kuvio T2.10 Kahden normaalijakauman sovittaminen sairastumisiin jakaumaan

Tiheysfunktioestimaatit [$\hat{f}(t)$ ML, $\hat{f}(t)$ ad hoc] verrattuna havaittuun frekvenssiin



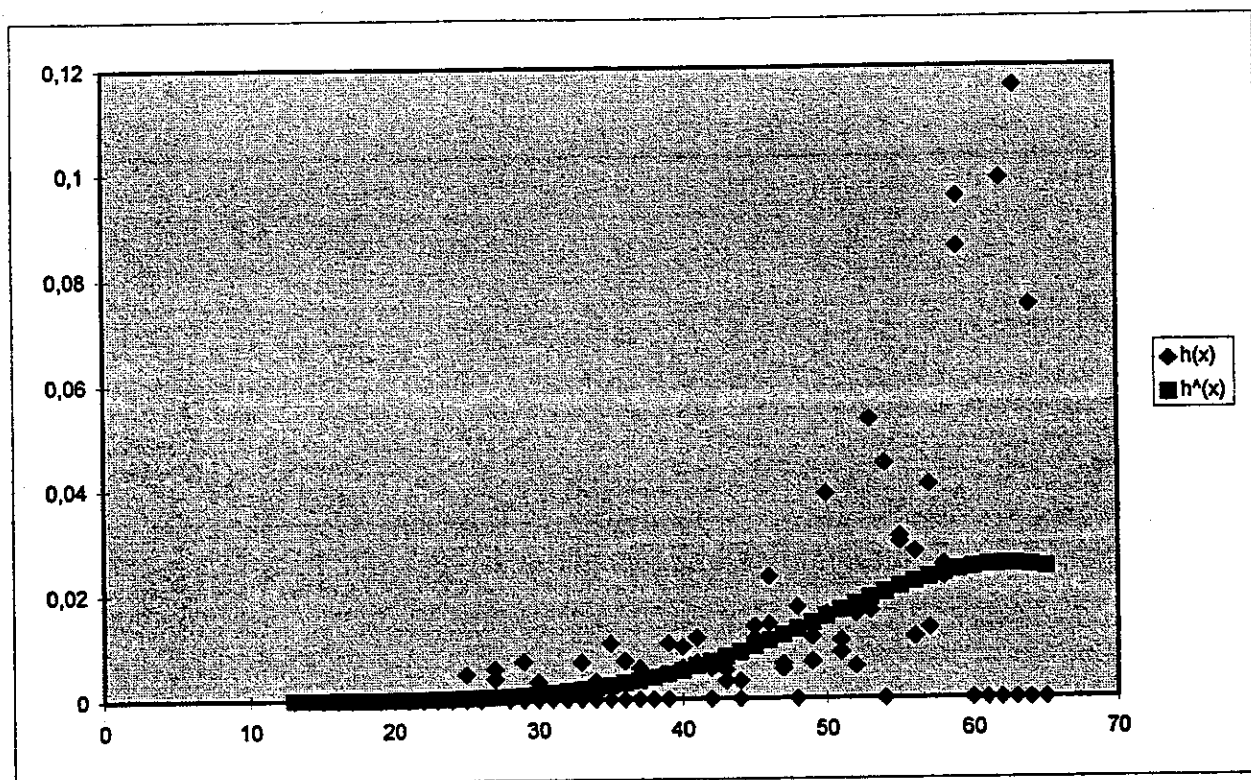
Kuvio T2.11 Log-logistisen jakauman sovittaminen sairastumisiin jakaumaan

Tiheysfunktioestimaatit [$\hat{f}(t)$ ML, $\hat{f}(t)$ ad hoc] verrattuna havaittuun frekvenssiin



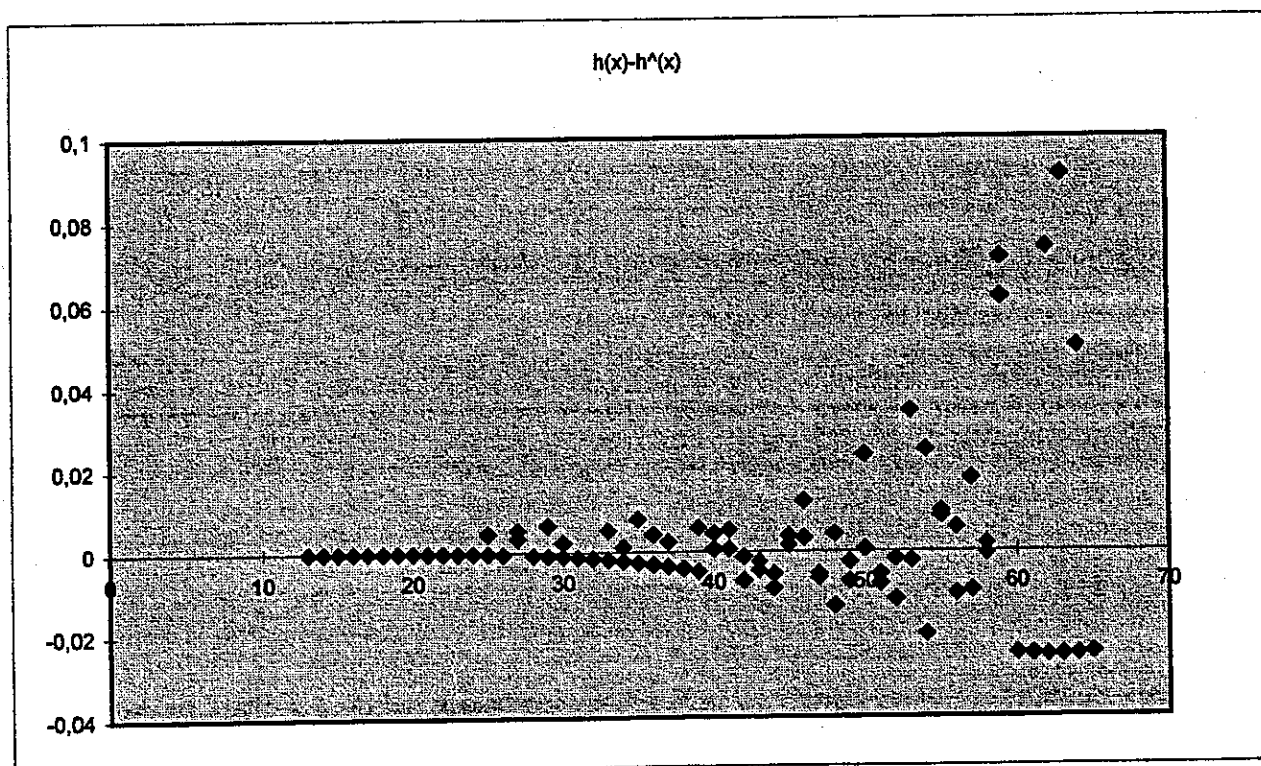
Liite T
Kuvio T2.12

Kahden normaalijakauman ad hoc sovituksen mukainen intensiteetti verrattuna havaittuihin.
Yksivuotisikäluokat, sukupuoli erikseen



Kuvio T2.13

'Residuaalikuvio' havaitun ja mallin mukaisen intensiteetin ero.
Yksivuotisikäluokat, sukupuoli erikseen.



Liite T

Taulukko T3.1 Käytössä olevat muuttujat

DD	seurannassaoloaika vuosina aikaväliltä 1.1.1994 -31.12.1996
S	parantumisen indikaattorimuuttuja (1=parantui, 0=kato)
Q	kuolemissen indikaattorimuuttuja (1=kuolema, 0=kato)
W = S+Q	yhdistetty indikaattorimuuttuja (1=tapahtuma, 0=kato)
SPN	sukupuoli (1=nainen, 0=mies)
TVJ = (TV-13)/13	TV = ikä eläketapahtumavuonna
SVJ = (SV-1930)/1930	SV = syntymävuosi
SVL1930	=1, jos SV on 1930-luvulla, muutoin = 0
SVL1940	=1, jos SV on 1940-luvulla, muutoin = 0
SVL1950	=1, jos SV on 1950-luvulla, muutoin = 0
SVL1960	=1, jos SV on 1960-luvulla, muutoin = 0
SVL1970	=1, jos SV on 1970-luvulla, muutoin = 0
TVL4	=1, jos TV on pienempi tai yhtäkuin 59, muutoin = 0, indikaattorimuuttuja ns. aidolle tk-tapaukselle
TVC = SPN * TVJ	kovarianssitermi naisten iälle tapahtumavuonna
SVC = SPN * SVJ	kovarianssitermi naisten syntymävuodelle
TVL500 = TVL4 * TVJ	kovarianssitermi ns. aidolle tk-tapaukselle ja iälle tapahtumavuonna

Liite T

Taulu T3.2 Jakaumasovituksen tulokset, SAS-tulosteet

Selitettävä log(DD), tapahtuma S

parametriestimaatit	Log-logistic	Weibull	Lognormaali
INTERCEPT	4,18572	4,21167	5,68738
SCALE	0,81542	0,81988	2,27791
-logL	3.034,67	3.034,67	3.038,76

Selitettävä log(DD), tapahtuma Q

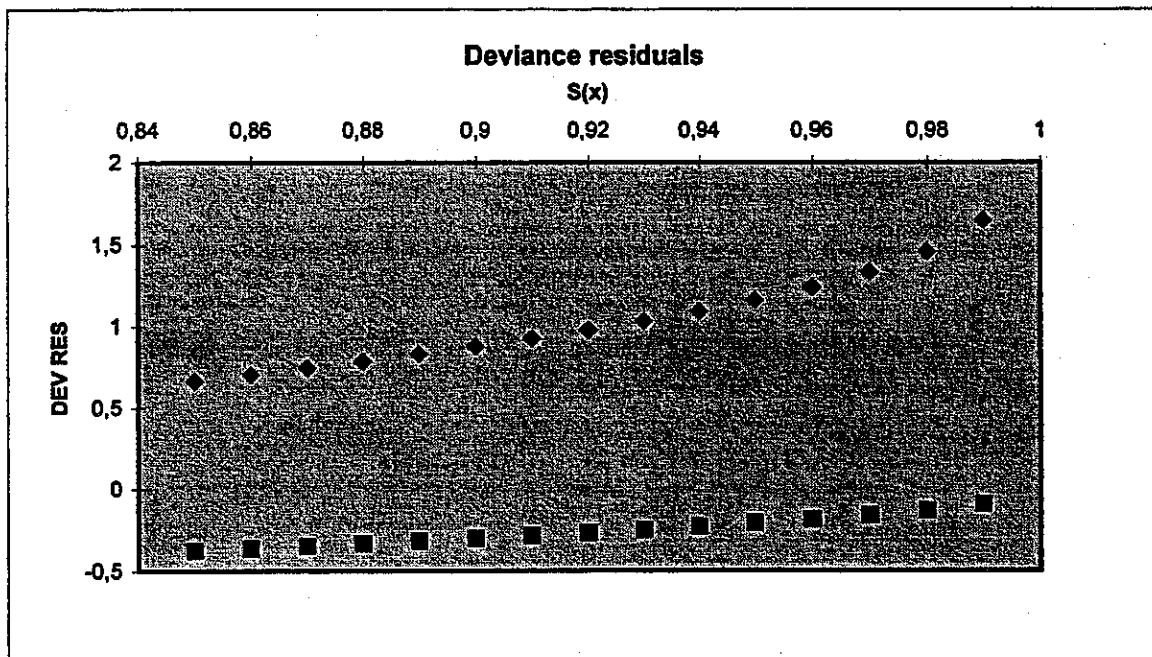
parametriestimaatit	Log-logistic	Weibull	Lognormaali
INTERCEPT	3,54871	3,60412	4,51134
SCALE	0,85104	0,86215	2,10870
-logL	6.327,96	6.328,22	6.327,20

Selitettävä log(DD), tapahtuma W

parametriestimaatit	Log-logistic	Weibull	Lognormaali
INTERCEPT	3,20475	3,27629	3,98491
SCALE	0,83441	0,84988	1,98988
-logL	8.254,44	8.254,74	8.260,86

Kuvio T3.10

Teoreettisten deviance residuaalien käyttäytyminen eri elossaolotodennäköisyyden arvoilla



Liite T Taulu T3.3 Valittujen mallien parametrit

Selitettävä DD, kato S

parametristimaatit	Log-logistic	riski	Weibull	riski	Cox-model	riski
λ	1,5422E-10		2,396E-10			
SVL1940	1,41161	4,103	1,41880	4,132	1,39779	4,046
SVL1950	1,45113	4,268	1,45743	4,295	1,45852	4,300
SVL1960	0,99190	2,696	0,98205	2,670	0,97072	2,640
TVJ	788,24023		765,35379		753,99820	
SVJ	4,01972	1,002	3,92166	1,002	3,85132	1,002
α	1,33181		1,30092			

Selitettävä DD, kato Q

parametristimaatit	Log-logistic	riski	Weibull	riski	Cox-model	riski
λ	0,06751731		0,0639443			
SPN	-1,71457	0,180	-1,66617	0,189	-1,675057	0,187
TVL4	-0,32945	0,719	-0,32048	0,726	-0,327071	0,721
TVJ	-0,22644	0,983	-0,21843	0,983	-0,216797	0,983
SVJ	-71,62127	0,964	-69,62877	0,965	-70,54491	0,964
TVC	0,28571	1,022	0,27632	1,021	0,270181	1,021
SVC	79,73492	1,042	77,48002	1,041	80,369918	1,043
α	1,18436		1,16774			

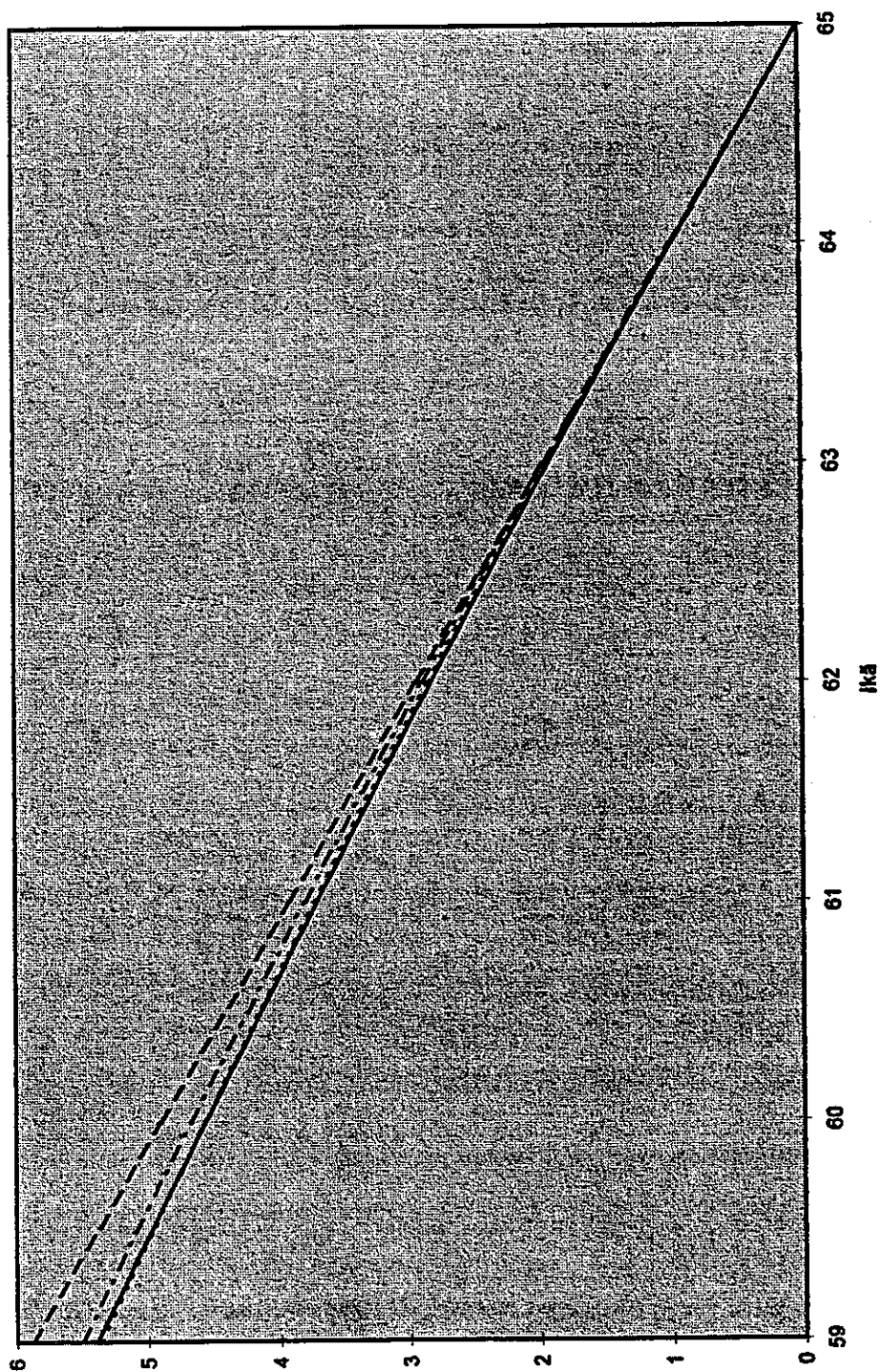
Selitettävä DD, kato W

parametristimaatit	Log-logistic	riski	Weibull	riski	Cox-model	riski
λ	0,01239345		0,011763			
SPN	-2,17348	0,114	-2,14502	0,117	-2,144545	0,117
SVL1930	-0,38564	0,680	-0,36607	0,693	-0,352897	0,703
SVL1940	-0,28176	0,754	-0,26655	0,766	-0,264153	0,768
TVJ	0,26088	1,020	0,26727	1,021	0,260923	1,020
SVJ	58,66481	1,031	59,34245	1,031	58,437273	1,031
TVC	0,42275	1,033	0,41930	1,033	0,41273	1,032
SVC	105,16732	1,056	103,77195	1,055	104,59477	1,056
α	1,20347		1,17725			

riski = yhden yksikön lisäyksen aiheuttama vaaran lisääntymä

Liite T Kuvio T3.1 Log-logistisen mallin ennuste odotusarvolle verrattuna nykyiseen perusteeseen sekä 13-päivänä syntyneiden havaittuun arvoon

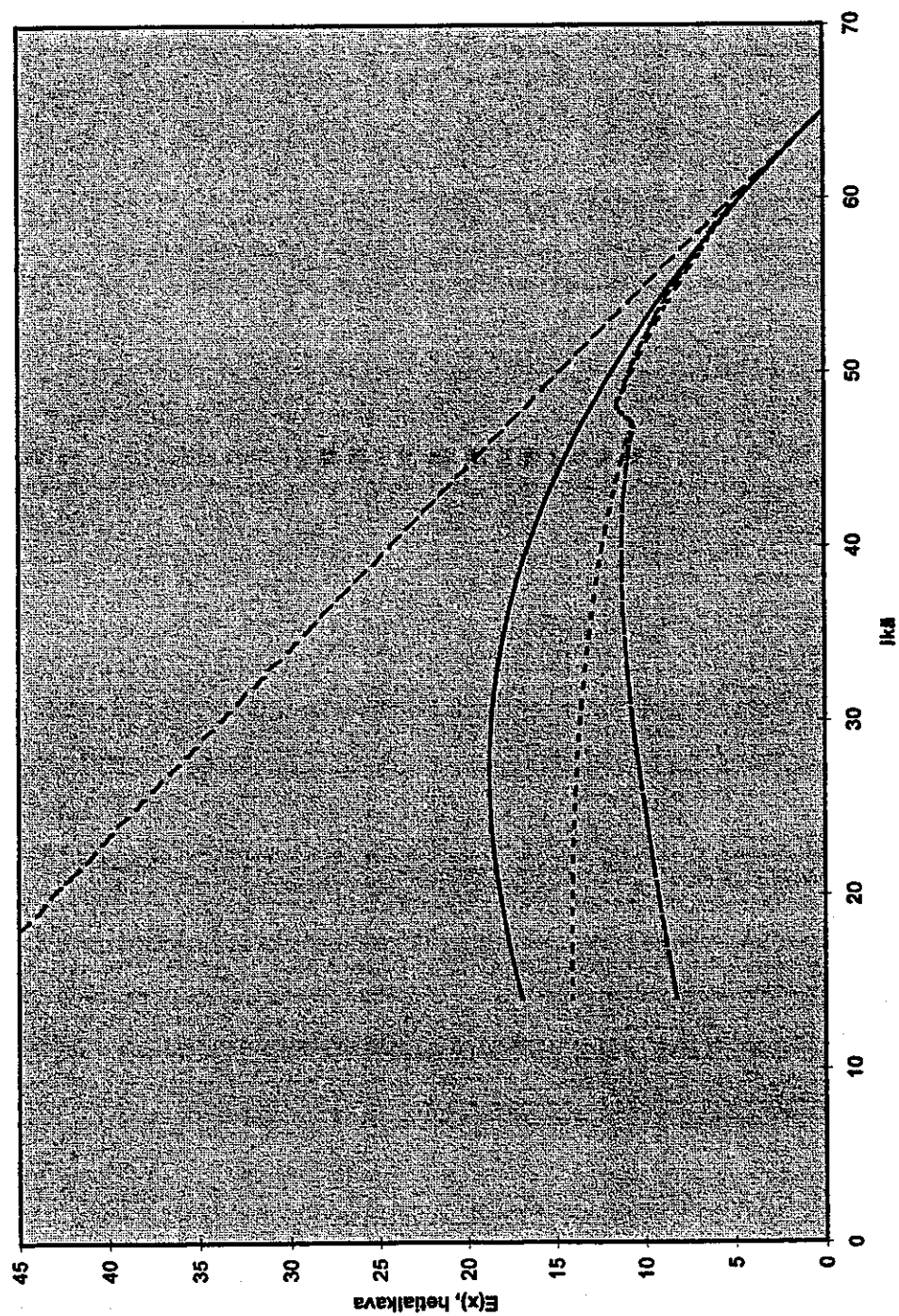
$E(x)$



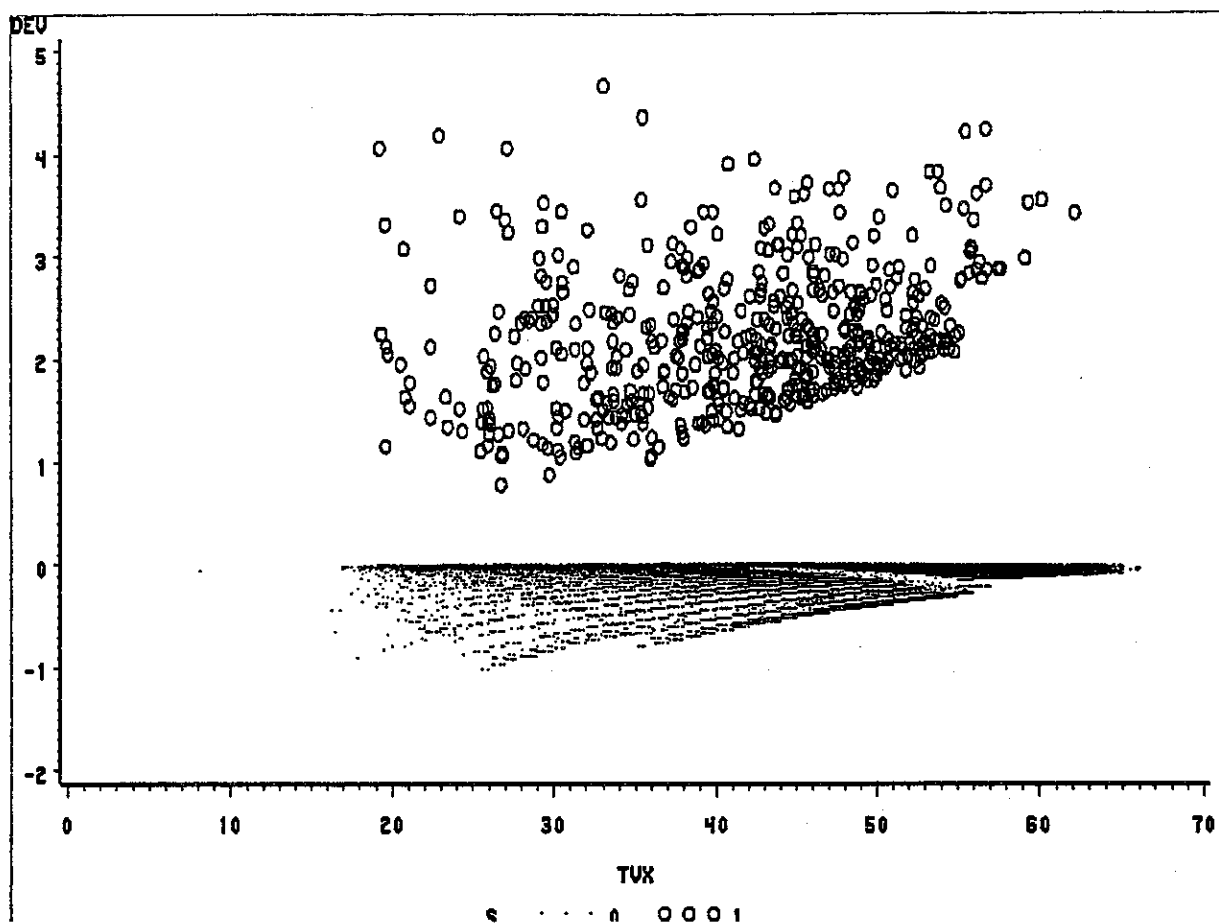
nykyinen peruste
 13-päivä syntyneet, tapa B
 MIES, UUSI
 NAINEN, UUSI

Liite T Kuvio T3.1b Log-logistisen mallin ennuste odotusarvolle verrattuna nykyiseen perusteeseen sekä 13-päivänä syntyneiden havaittuun arvoon

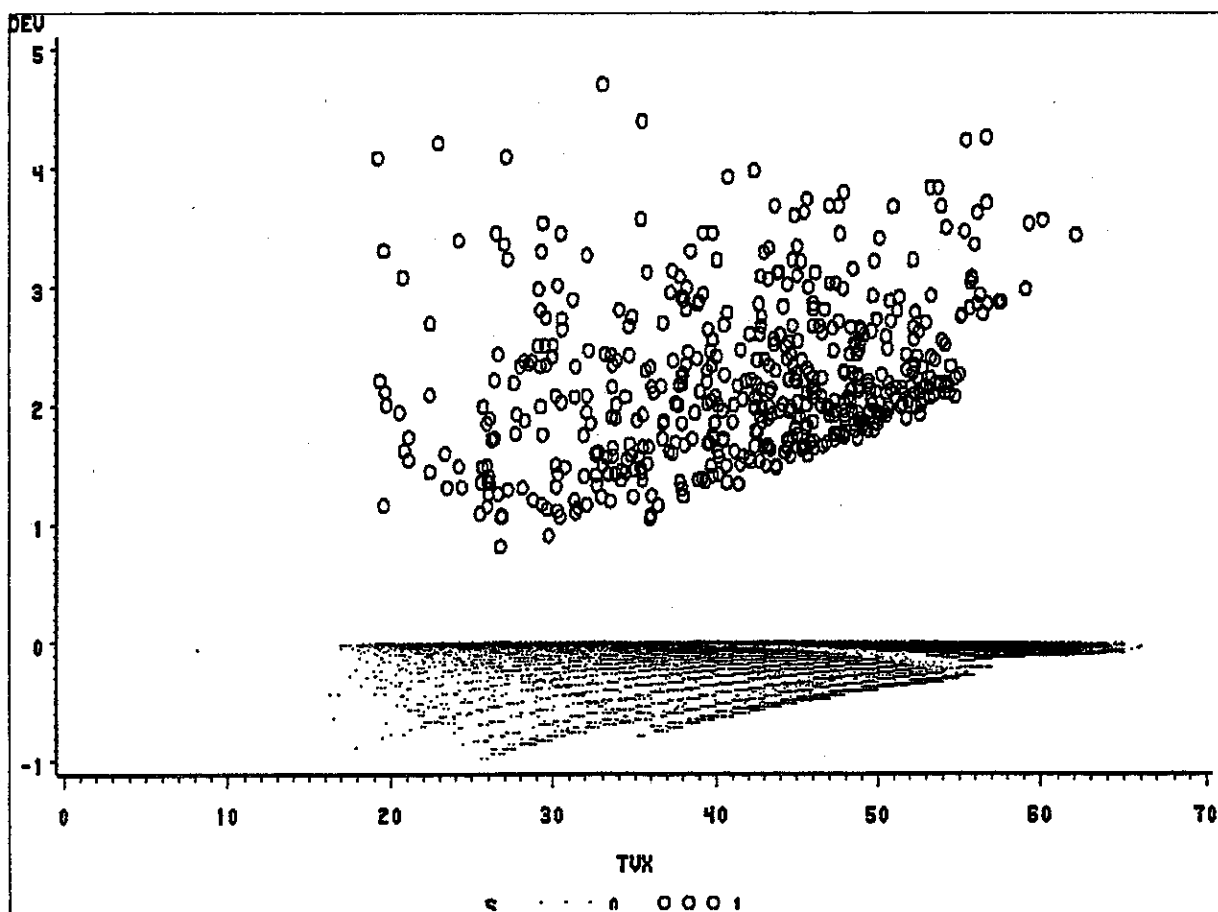
Havainnot, uusi malli ja vanha peruste



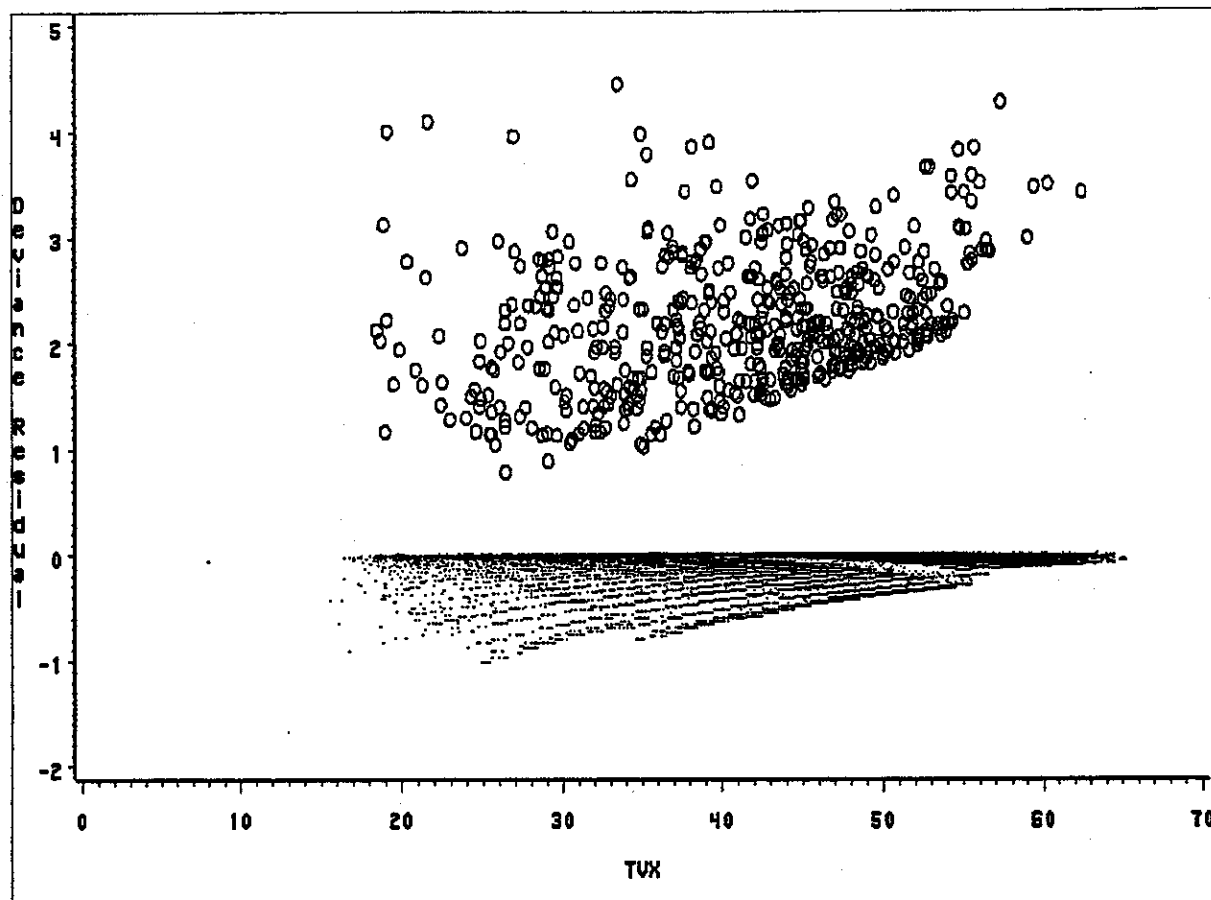
Kuvio T3.2 Työkyvyttömyyseläkkeeltä parantuneet, Weibull mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



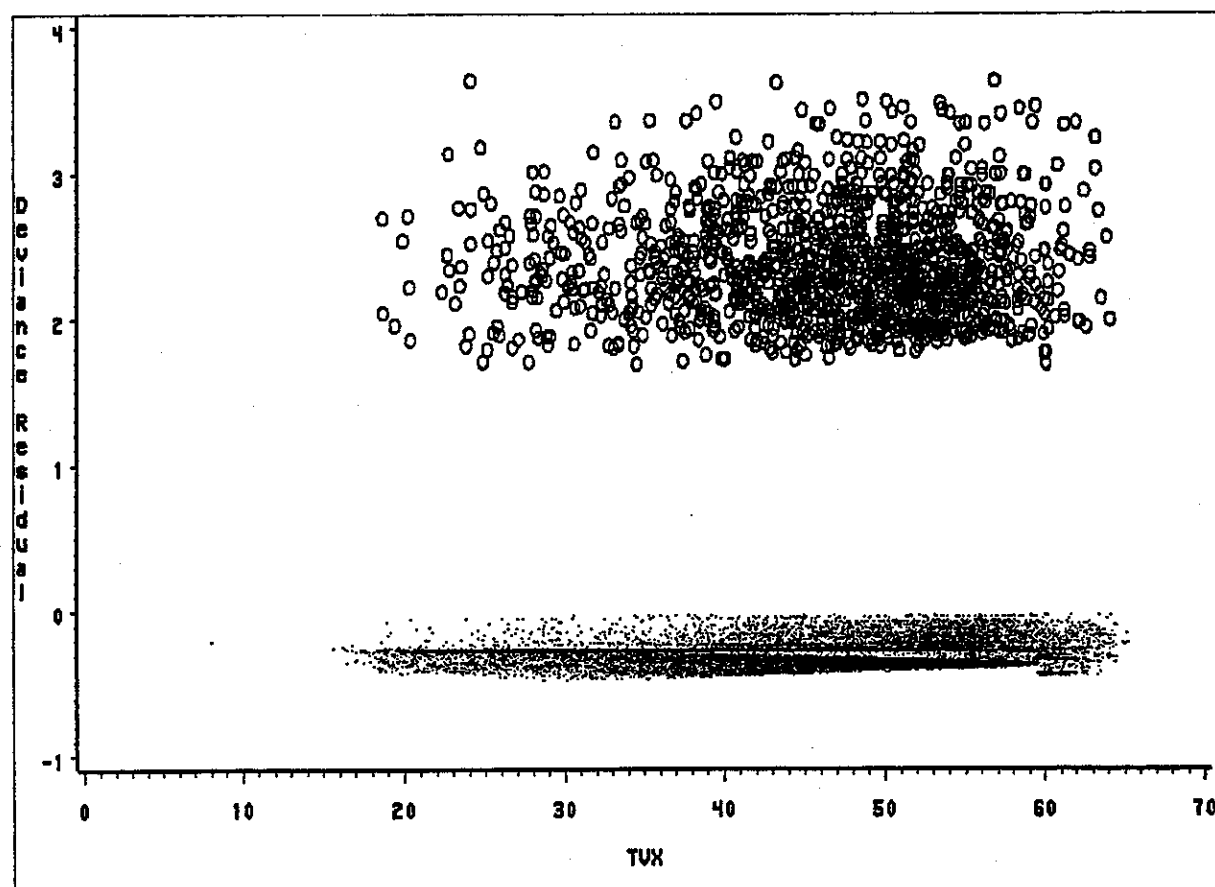
Kuvio T3.3 Työkyvyttömyyseläkkeeltä parantuneet, Log-logistisen mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



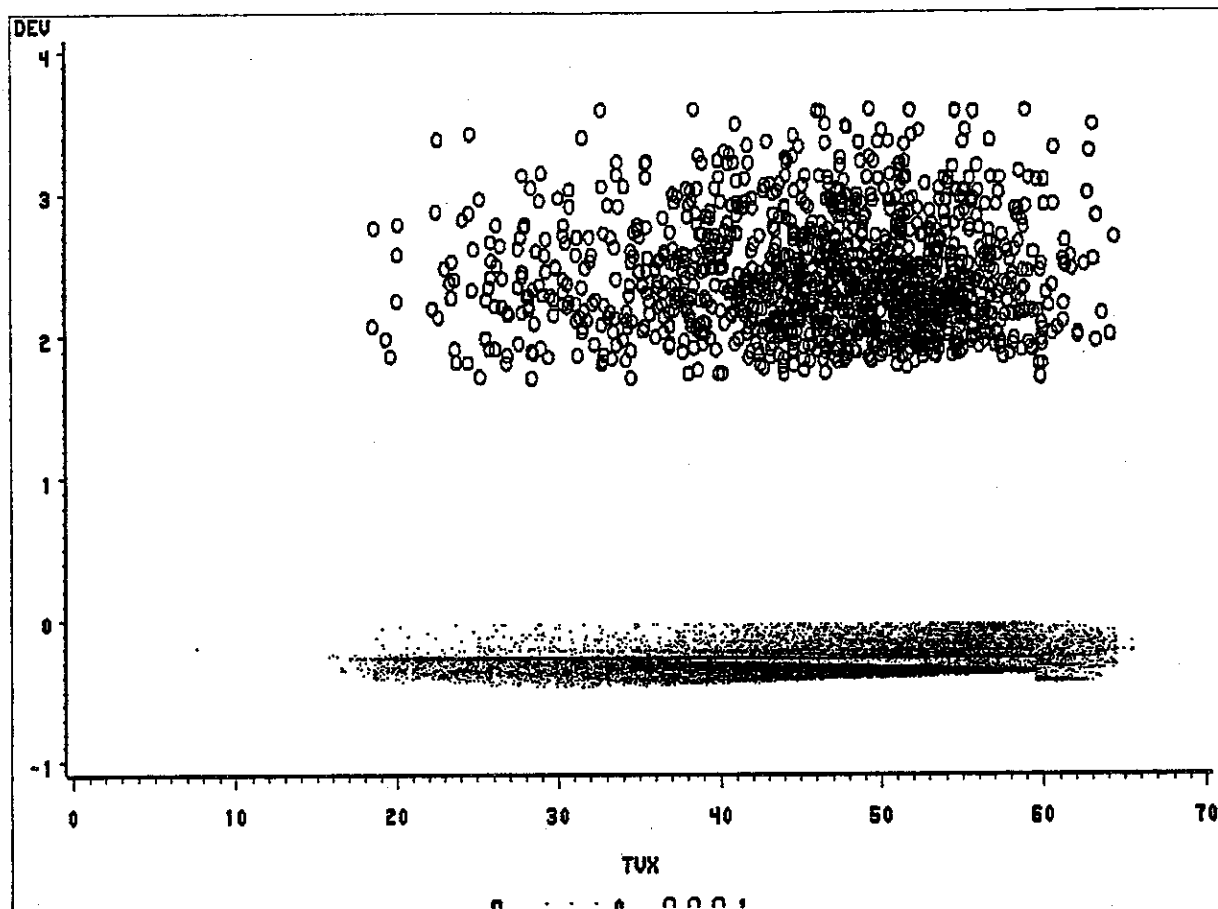
Kuvio T3.4 Työkyvyttömyyseläkkeeltä parantuneet, Coxin mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



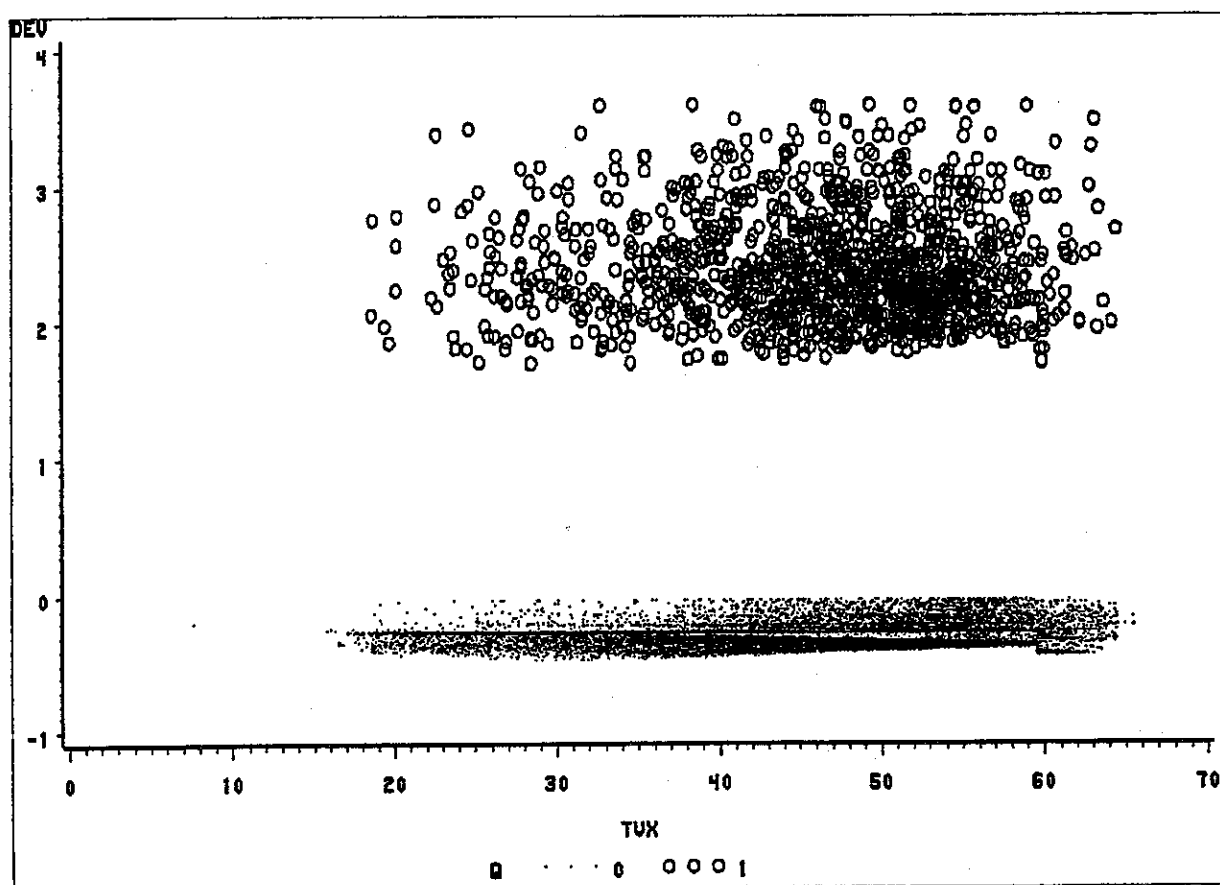
Kuvio T3.5 Työkyvyttömyyseläkkeeltä kuolleet, Coxin mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



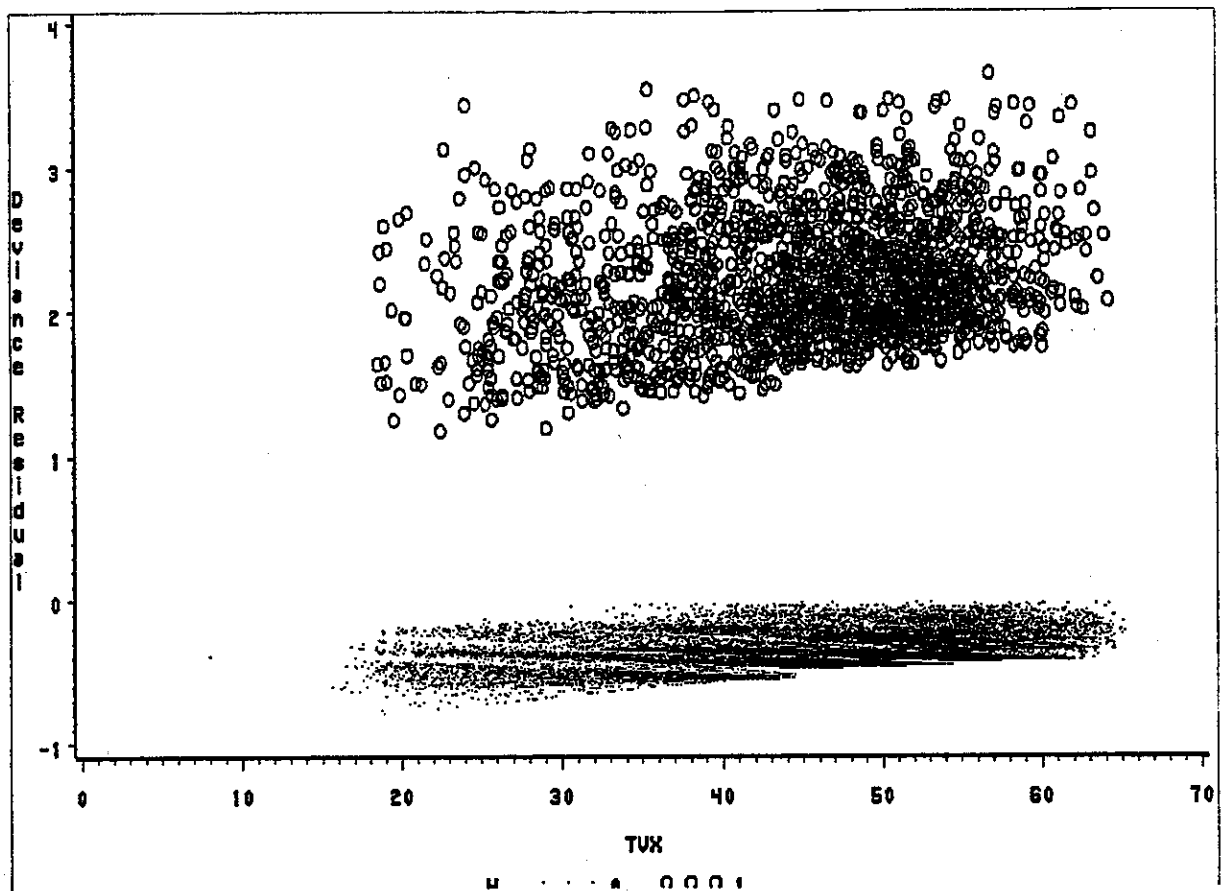
Kuvio T3.6 Työkyvyttömyyseläkkeeltä kuolleet, Weibull mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



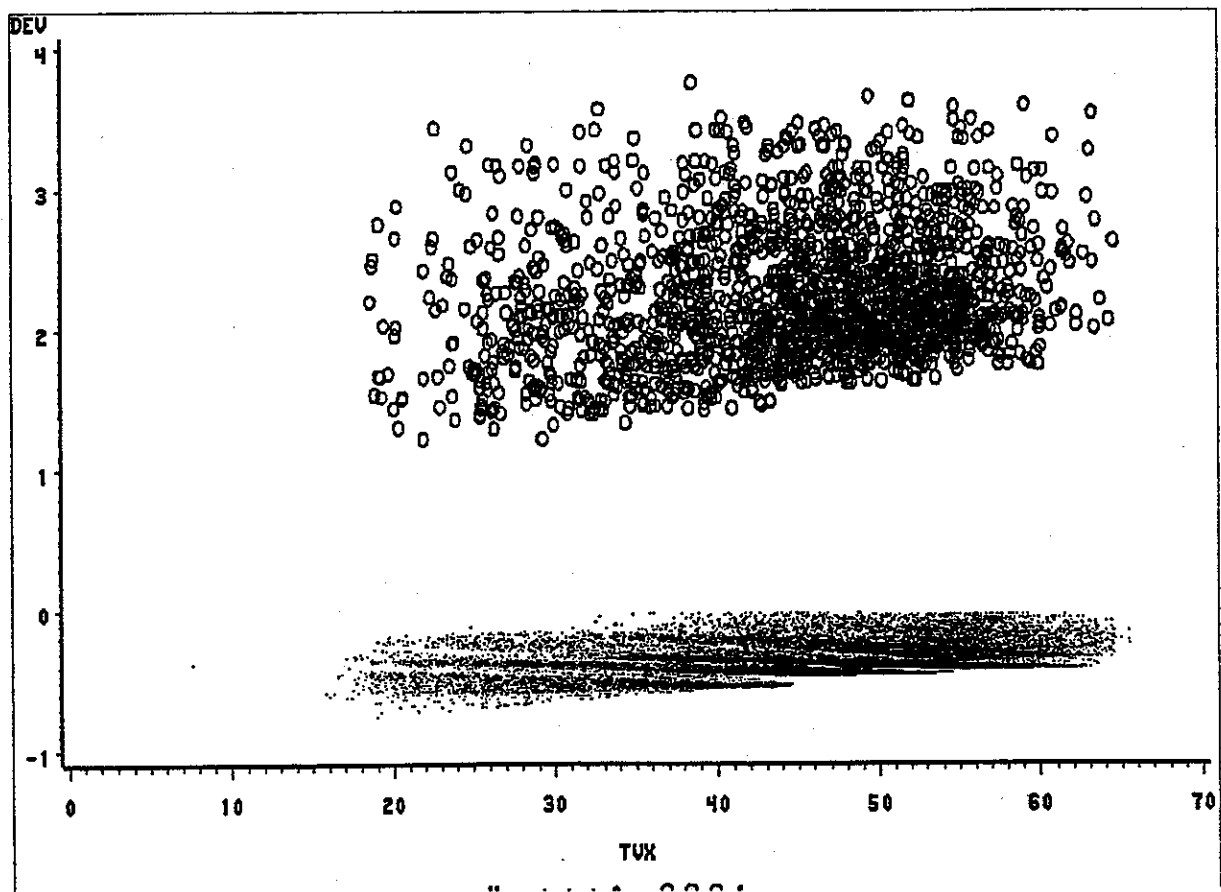
Kuvio T3.7 Työkyvyttömyyseläkkeeltä kuolleet, log-logistisen mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



Kuvio T3.8 Työkyvyttömyyseläkkeeltä poistuneet, Coxin mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



Kuvio T3.9 Työkyvyttömyyseläkkeeltä poistuneet, log-logistisen mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)

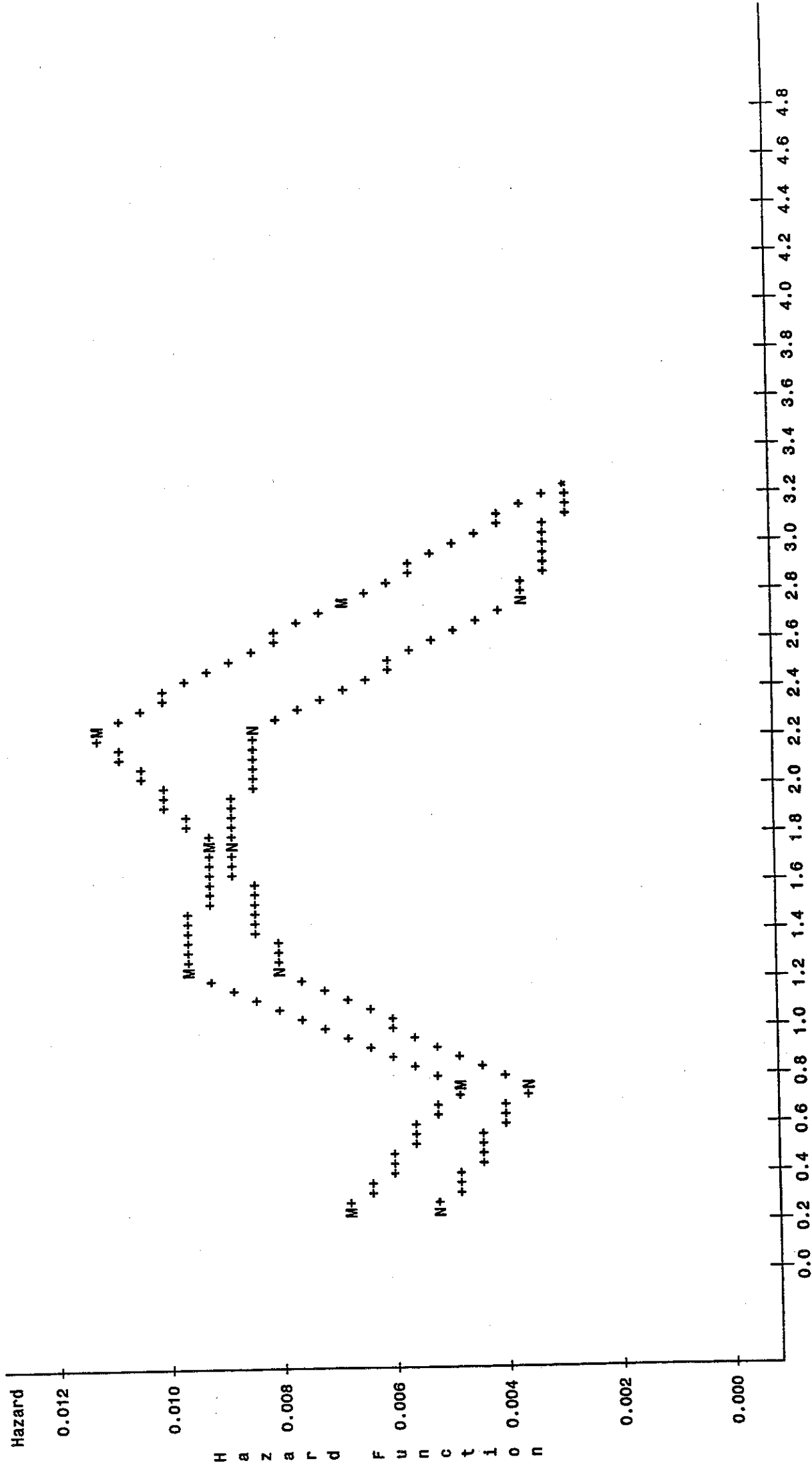


White S Kunz73.1
Parantuneet

The SAS System

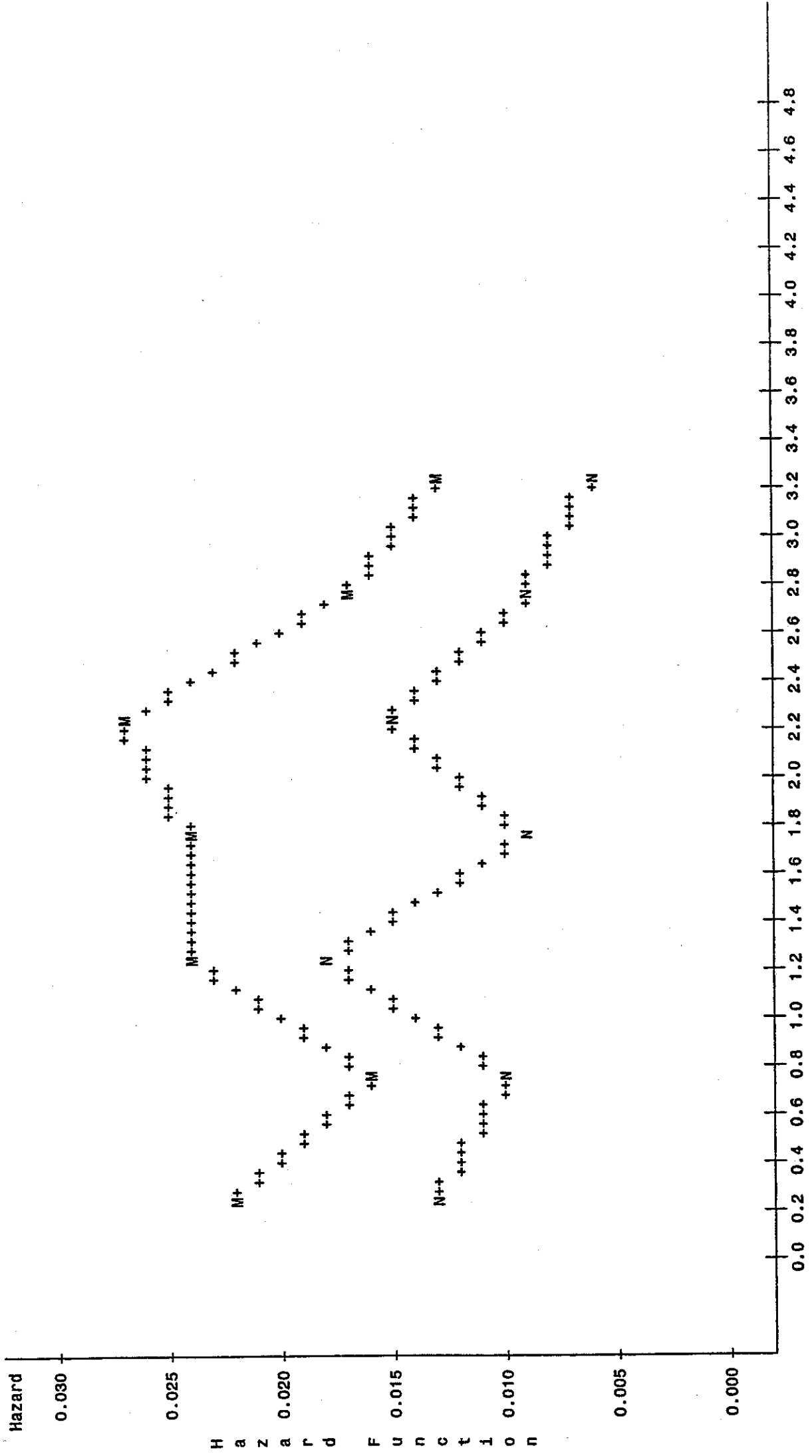
The LIFETEST Procedure

Hazard Function Estimates



The LIFETEST Procedure

Hazard Function Estimates

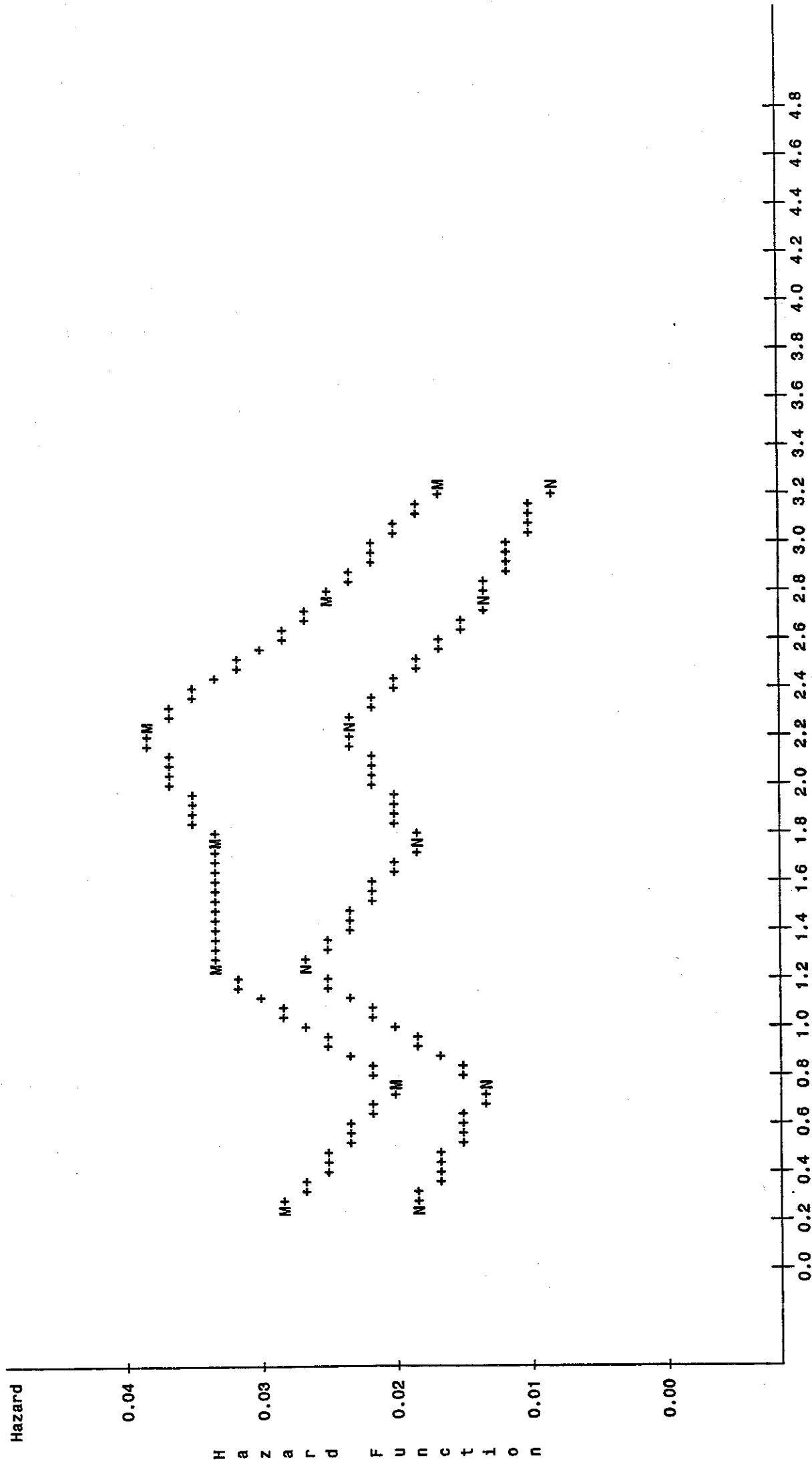


88

Write S Kurv T3.3
TK-eläke päättyyt

The LIFETEST Procedure

Hazard Function Estimates



The SAS System

The PHREG Procedure

Data Set: WORK.DL4
 Dependent Variable: DD
 Censoring Variable: S
 Censoring Value(s): 0
 Ties Handling: BRESLOW

Summary of the Number of
 Event and Censored Values

Total	Percent	
	Event	Censored
27828	540	27288
		98.06

Testing Global Null Hypothesis: BETA=0

Criterion	Without		With		Model Chi-Square
	Covariates		Covariates		
-2 LOG L	10897.075		9481.626		1415.449 with 5 DF (p=0.0001)
Score	.		.		1454.222 with 5 DF (p=0.0001)
Wald	.		.		769.210 with 5 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald		Pr > Chi-Square	Risk Ratio
				Chi-Square			
SVL1940	1	1.397789	0.18214	58.89728		0.0001	4.046
SVL1950	1	1.458519	0.20480	50.72012		0.0001	4.300
SVL1960	1	0.970722	0.26613	13.30462		0.0003	2.640
SVJ	1	753.998198	37.20289	410.75918		0.0001	.
TVJ	1	3.851324	0.22891	283.06404		0.0001	47.055

Lifereg Procedure

Data Set =WORK.DL4
Dependent Variable=Log(DD)
Censoring Variable=S
Censoring Value(s)= 0
Noncensored Values= 534 Right Censored Values= 2737
Left Censored Values= 0 Interval Censored Values= 0
Observations with Zero or Negative Response= 57

Log Likelihood for LLOGLSTC -2328.435332

The SAS System

Lifereg Procedure

Variable DF Estimate Std Err ChiSquare Pr>Chi Label/Value

INTERCPT	1	16.9638038	0.897685	357.1073	0.0001	Intercept
SVL1940	1	-1.0599124	0.148715	50.79603	0.0001	
SVL1950	1	-1.0895898	0.168653	41.73857	0.0001	
SVL1960	1	-0.7477251	0.21574	12.01216	0.0005	
SVJ	1	-591.85431	36.22764	266.9005	0.0001	
TVJ	1	-3.0182283	0.205697	215.3012	0.0001	
SCALE	1	0.75085524	0.031098			Logistic scale parameter

Data Set =WORK.DL4
 Dependent Variable=Log(DD)
 Censoring Variable=S
 Censoring Value(s)= 0
 Noncensored Values= 534 Right Censored Values= 27237
 Left Censored Values= 0 Interval Censored Values= 0
 Observations with Zero or Negative Response= 57

Log Likelihood for WEIBULL -2326.364307

The SAS System

Lifereg Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	17.0280857	0.8973	360.1273	0.0001	Intercept
SVL1940	1	-1.0906164	0.148993	53.58146	0.0001	
SVL1950	1	-1.1203086	0.165325	45.91843	0.0001	
SVL1960	1	-0.7584888	0.207799	13.32334	0.0003	
SVJ	1	-586.31783	35.78457	270.2915	0.0001	
TVJ	1	-3.0145341	0.20479	216.6821	0.0001	
SCALE	1	0.76868742	0.032005			Extreme value scale parameter

The PHREG Procedure

Data Set: WORK.DL4
 Dependent Variable: DD
 Censoring Variable: Q
 Censoring Value(s): 0
 Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
27828	1327	26501	95.23

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	26792.157	26652.187	139.970 with 6 DF (p=0.0001)
Score	.	.	144.610 with 6 DF (p=0.0001)
Wald	.	.	139.887 with 6 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
SPN	1	-1.675067	0.39739	17.76701	0.0001	0.187
TVL4	1	-0.327071	0.15867	4.24884	0.0393	0.721
TVJ	1	-0.216797	0.06528	11.03451	0.0009	0.805
SVJ	1	-70.544910	11.89166	35.19219	0.0001	0.000
TVC	1	0.270181	0.11976	5.09005	0.0241	1.310
SVC	1	80.369918	21.32556	14.20320	0.0002	8.021E34

The SAS System

Lifereg Procedure

Data Set =WORK.DL4
Dependent Variable=Log(DD)
Censoring Variable=Q
Censoring Value(s)= 0
Noncensored Values= 1307 Right Censored Values= 26464
Left Censored Values= 0 Interval Censored Values= 0
Observations with Zero or Negative Response= 57

Log Likelihood for LLOGISTIC -6261.875015

The SAS System

Lifereg Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	2.27561348	0.252453	81.2665	0.0001	Intercept
SPN	1	1.44767996	0.346924	17.41316	0.0001	
TVL4	1	0.27617006	0.140989	3.892671	0.0485	
TVJ	1	0.19118976	0.057767	10.9541	0.0009	
SVJ	1	60.4727987	10.51195	33.09434	0.0001	
TVC	1	-0.2412355	0.104141	5.365859	0.0205	
SVC	1	-67.323491	18.60779	13.09011	0.0003	
SCALE	1	0.84434138	0.022377			Logistic scale parameter

Lifereg Procedure

Data Set =WORK.DL4
 Dependent Variable=Log(DD)
 Censoring Variable=Q
 Censoring Value(s)= 0
 Noncensored Values= 1307 Right Censored Values= 26464
 Left Censored Values= 0 Interval Censored Values= 0
 Observations with Zero or Negative Response= 57

Log Likelihood for WEIBULL -6261.910887

The SAS System

Lifereg Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	2.35475366	0.248068	90.1054	0.0001	Intercept
SPN	1	1.42683531	0.344801	17.12424	0.0001	
TVL4	1	0.27444437	0.138646	3.918275	0.0478	
TVJ	1	0.18705159	0.056597	10.92295	0.0009	
SVJ	1	59.6268869	10.3698	33.06312	0.0001	
TVC	1	-0.2366238	0.103431	5.23377	0.0222	
SVC	1	-66.350331	18.52252	12.83176	0.0003	
SCALE	1	0.85635413	0.022833			Extreme value scale parameter

The SAS System

The PHREG Procedure

Data Set: WORK.DL4
 Dependent Variable: DD
 Censoring Variable: W
 Censoring Value(s): 0
 Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

	Total	Percent	
		Event	Censored
	27828	1867	25961
			93.29

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates		With Covariates		Model Chi-Square
-2 LOG L	37689.231		37322.619		366.612 with 7 DF (p=0.0001)
Score	.		.		355.976 with 7 DF (p=0.0001)
Wald	.		.		334.756 with 7 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
SPN	1	-2.144545	0.40098	28.60377	0.0001	0.117
SVL1930	1	-0.352897	0.12595	7.85090	0.0051	0.703
SVL1940	1	-0.264153	0.08851	8.90633	0.0028	0.768
TVJ	1	0.260923	0.06256	17.39288	0.0001	1.298
SVJ	1	58.437265	13.28536	19.34786	0.0001	2.393E25
TVC	1	0.412730	0.11667	12.51370	0.0004	1.511
SVC	1	104.594772	18.29243	32.69469	0.0001	2.66E45

Lifereg Procedure

Data Set =WORK.DL4
Dependent Variable=Log(DD)
Censoring Variable=W
Censoring Value(s)= 0
Noncensored Values= 1841 Right Censored Values= 25930
Left Censored Values= 0 Interval Censored Values= 0
Observations with Zero or Negative Response= 57

Log Likelihood for LLOGISTIC -8077.272294

The SAS System

Lifereg Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	3.64826122	0.225378	262.0281	0.0001	Intercept
SPN	1	1.80600743	0.344554	27.47418	0.0001	
SVL1930	1	0.32044139	0.111616	8.242218	0.0041	
SVL1940	1	0.23412199	0.078117	8.982471	0.0027	
TVJ	1	-0.2167684	0.054375	15.89257	0.0001	
SVJ	1	-48.746228	11.82351	16.99767	0.0001	
TVC	1	-0.3512752	0.100141	12.30459	0.0005	
SVC	1	-87.386455	15.85014	30.3984	0.0001	
SCALE	1	0.83092784	0.018623			Logistic scale parameter

Lifereg Procedure

Data Set =WORK.DL4
 Dependent Variable=Log(DD)
 Censoring Variable=W
 Censoring Value(s)= 0
 Noncensored Values= 1841 Right Censored Values= 25930
 Left Censored Values= 0 Interval Censored Values= 0
 Observations with Zero or Negative Response= 57

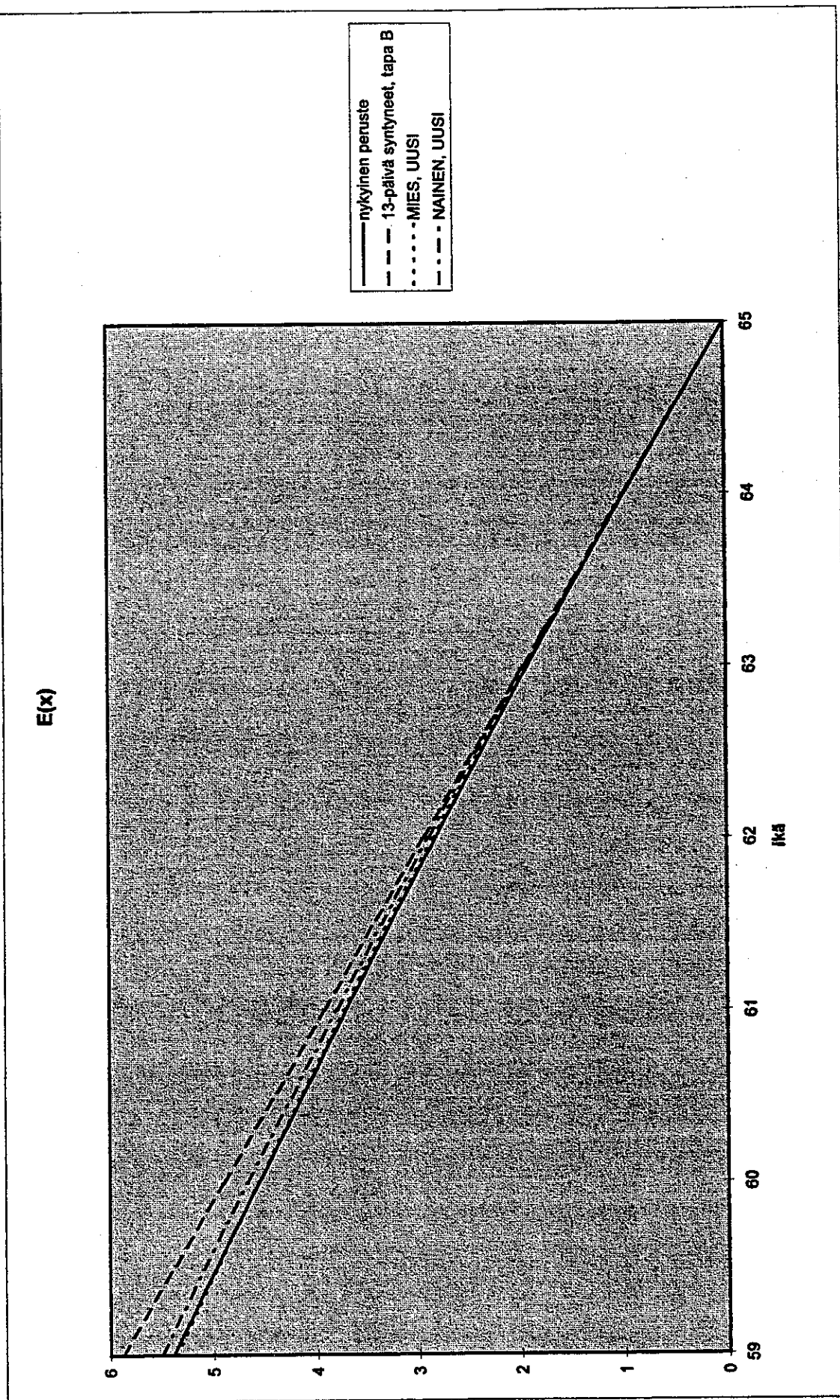
Log Likelihood for WEIBULL -8074.119763

The SAS System

Lifereg Procedure

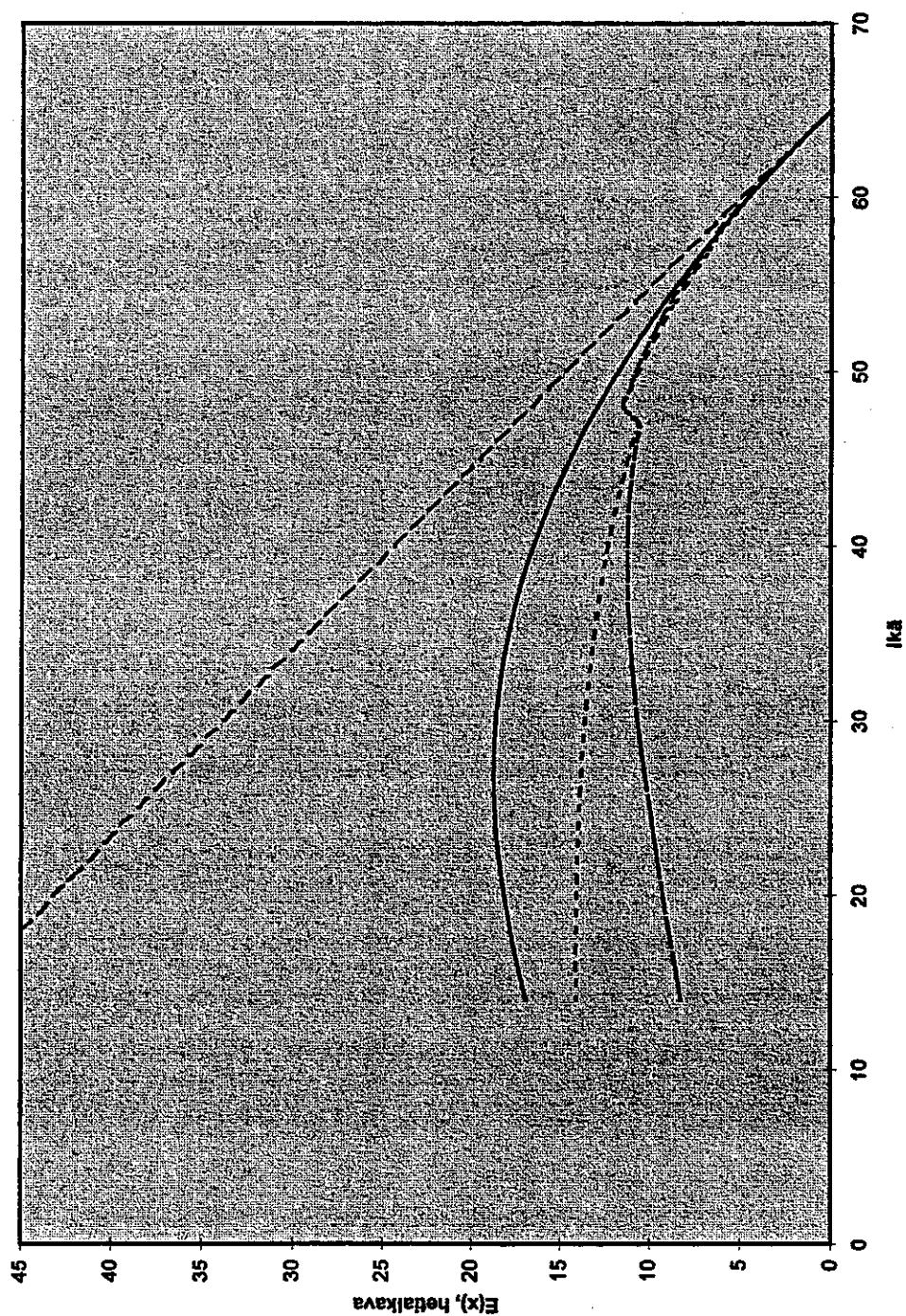
Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	3.77388042	0.223861	284.1972	0.0001	Intercept
SPN	1	1.82205997	0.345881	27.75058	0.0001	
SVL1930	1	0.31019578	0.109771	7.98534	0.0047	
SVL1940	1	0.22641337	0.07675	8.702628	0.0032	
TVJ	1	-0.2270316	0.053839	17.78209	0.0001	
SVJ	1	-50.407709	11.67012	18.65708	0.0001	
TVC	1	-0.3561723	0.100213	12.63208	0.0004	
SVC	1	-88.147801	15.80168	31.11833	0.0001	
SCALE	1	0.84943765	0.019208			Extreme value scale parameter

Liite T Kuvio T3.1 Log-logistisen mallin ennuste odotusarvolle verrattuna nykyiseen perusteeseen sekä 13-päivänä syntyneiden havaittuun arvoon

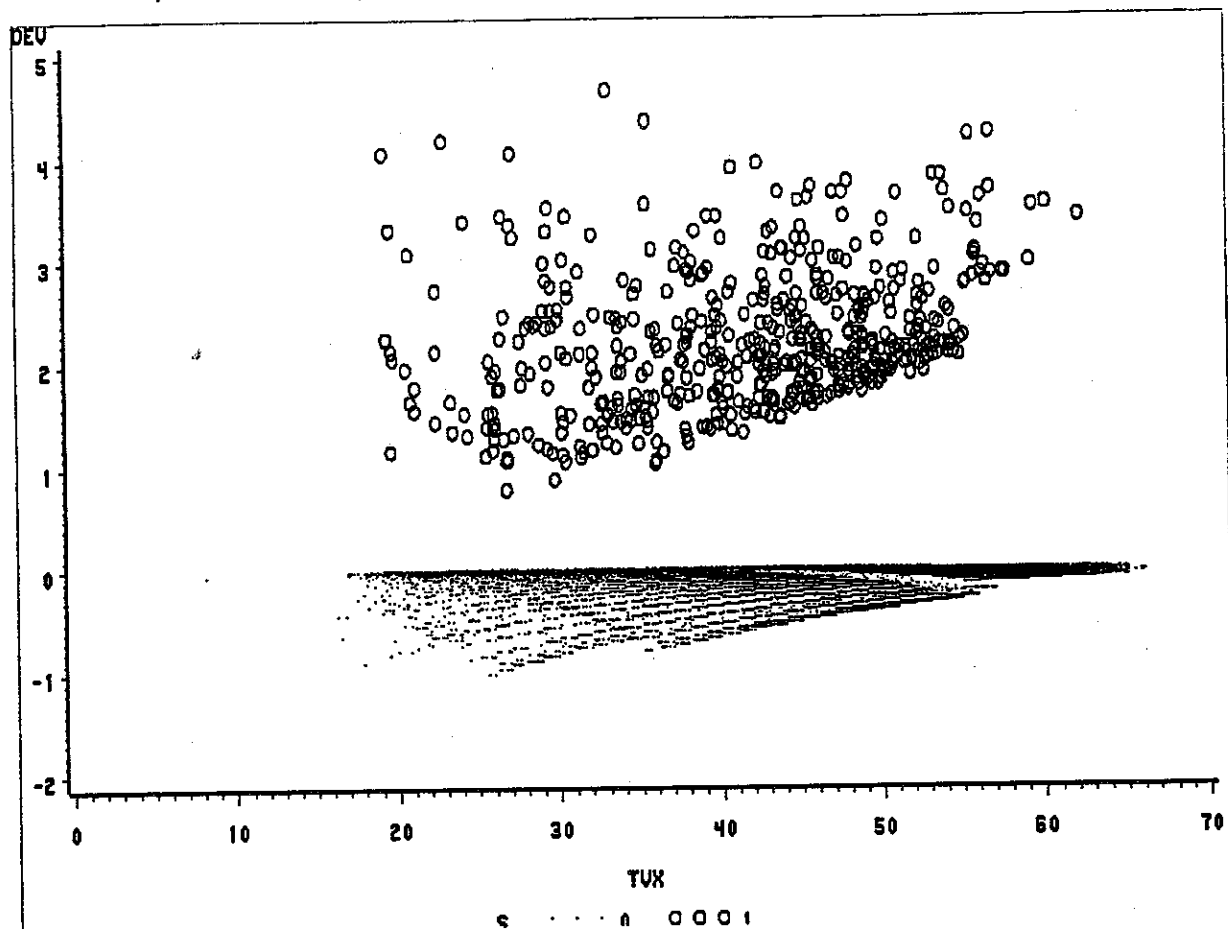


Liite T Kuvio T3.1b Log-logistisen mallin ennuste odotusarvolle verrattuna nykyiseen perusteeseen sekä 13-päivänä syntyneiden havaittuun arvoon

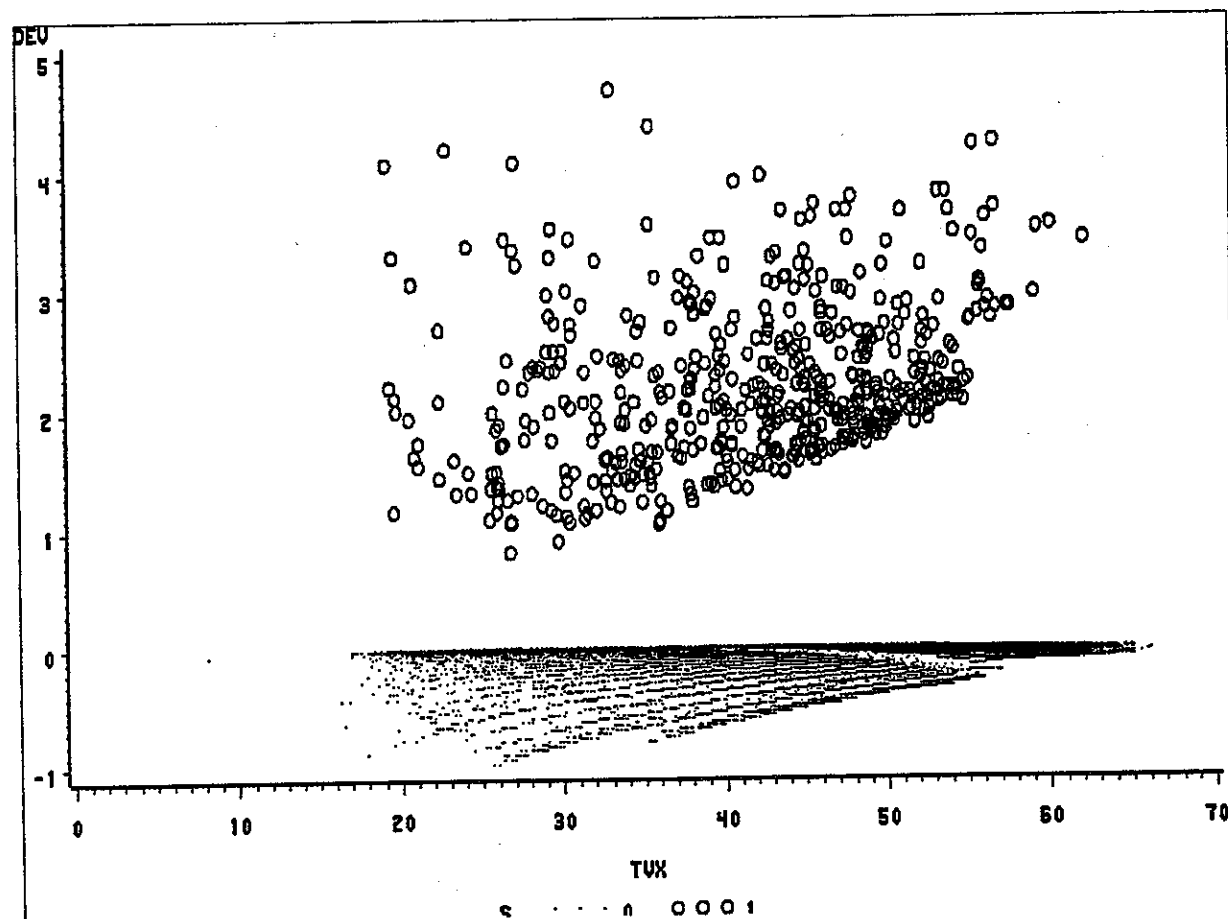
Havainnot, uusi malli ja vanha peruste



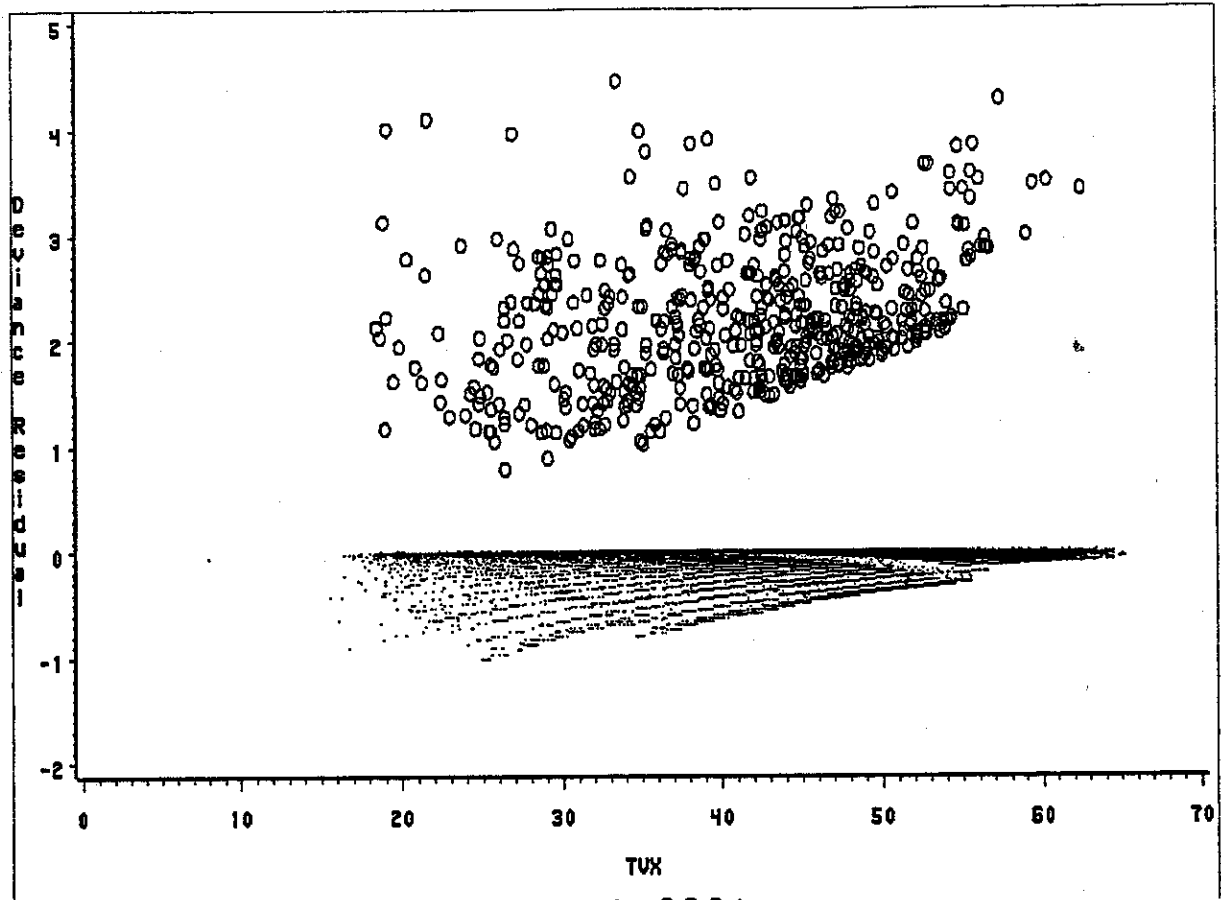
Kuvio T3.2 Työkyvyttömyyseläkkeeltä parantuneet, Weibull mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



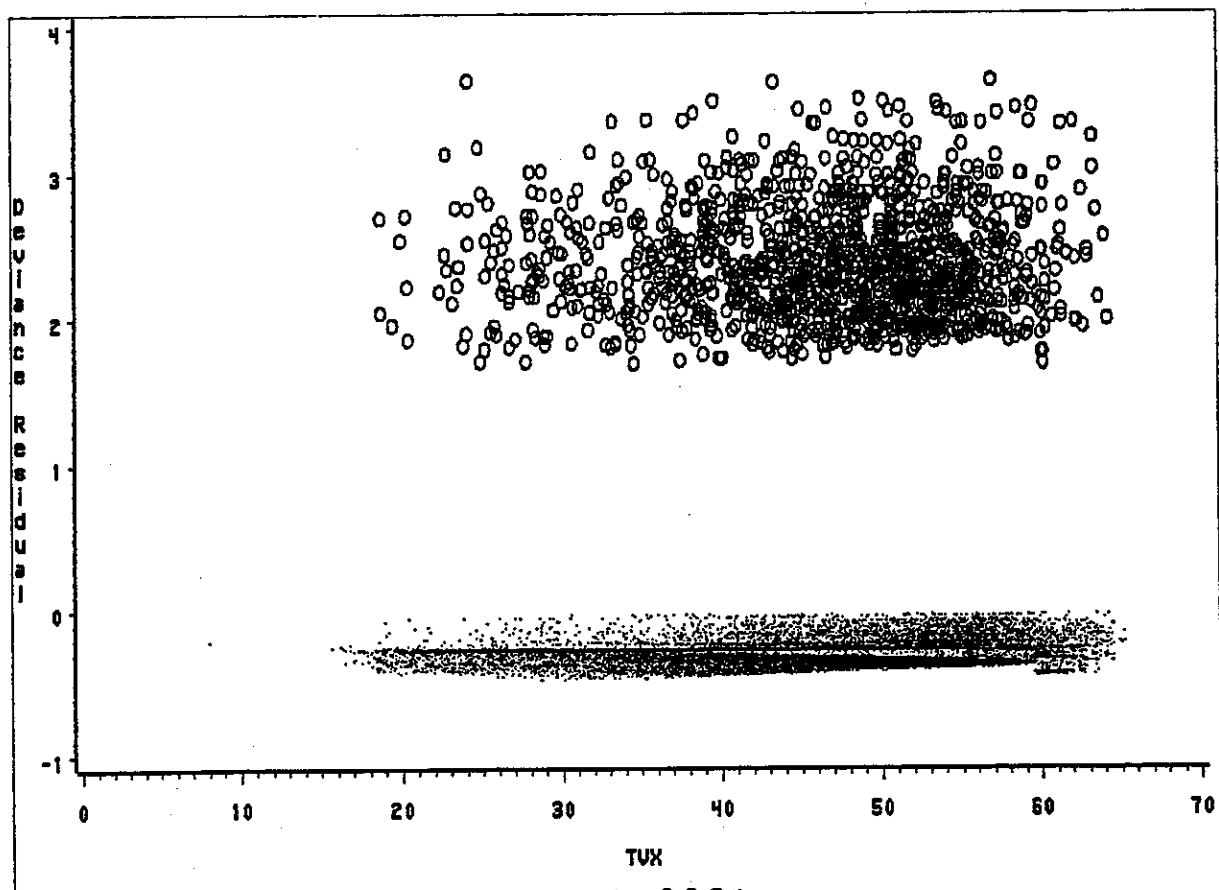
Kuvio T3.3 Työkyvyttömyyseläkkeeltä parantuneet, Log-logistisen mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



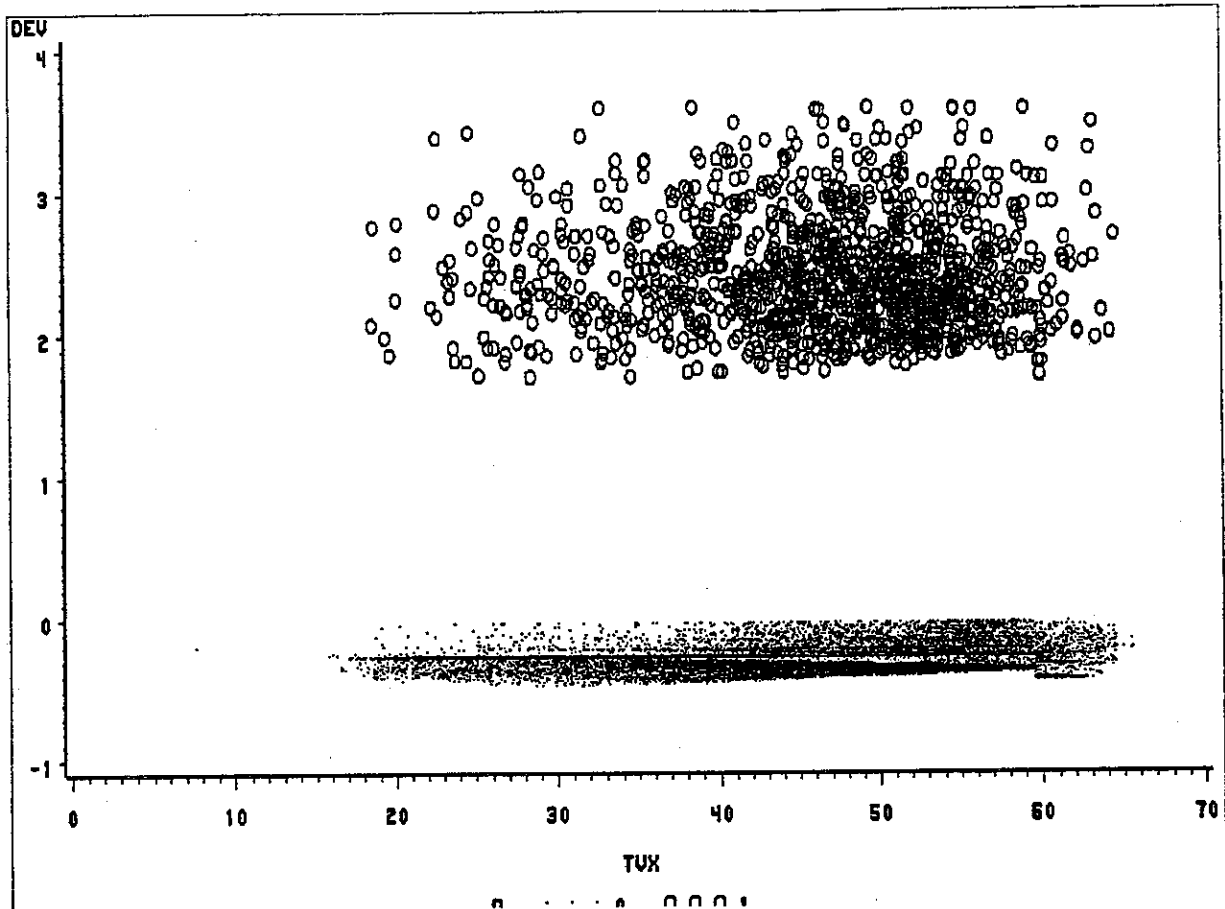
Kuvio T3.4 Työkyvyttömyyseläkkeeltä parantuneet, Coxin mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



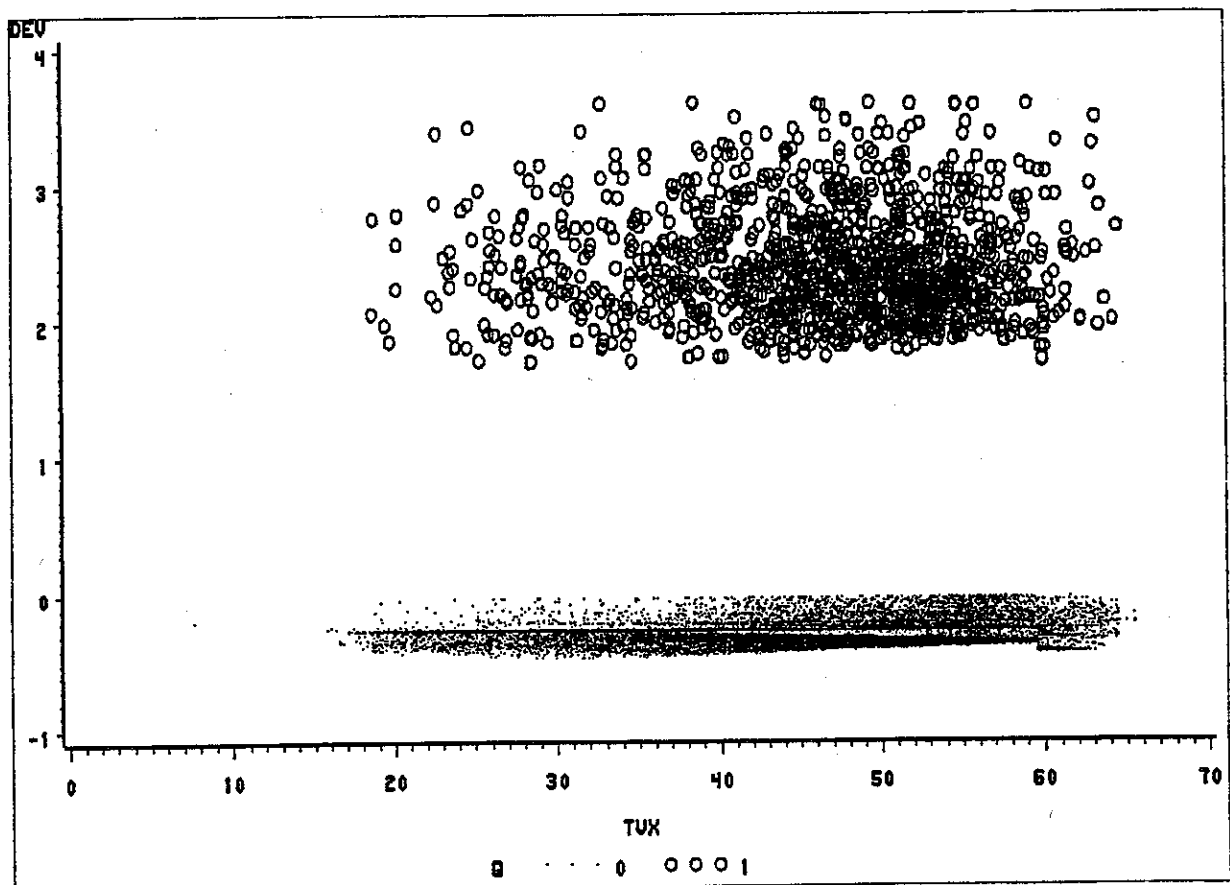
Kuvio T3.5 Työkyvyttömyyseläkkeeltä kuolleet, Coxin mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



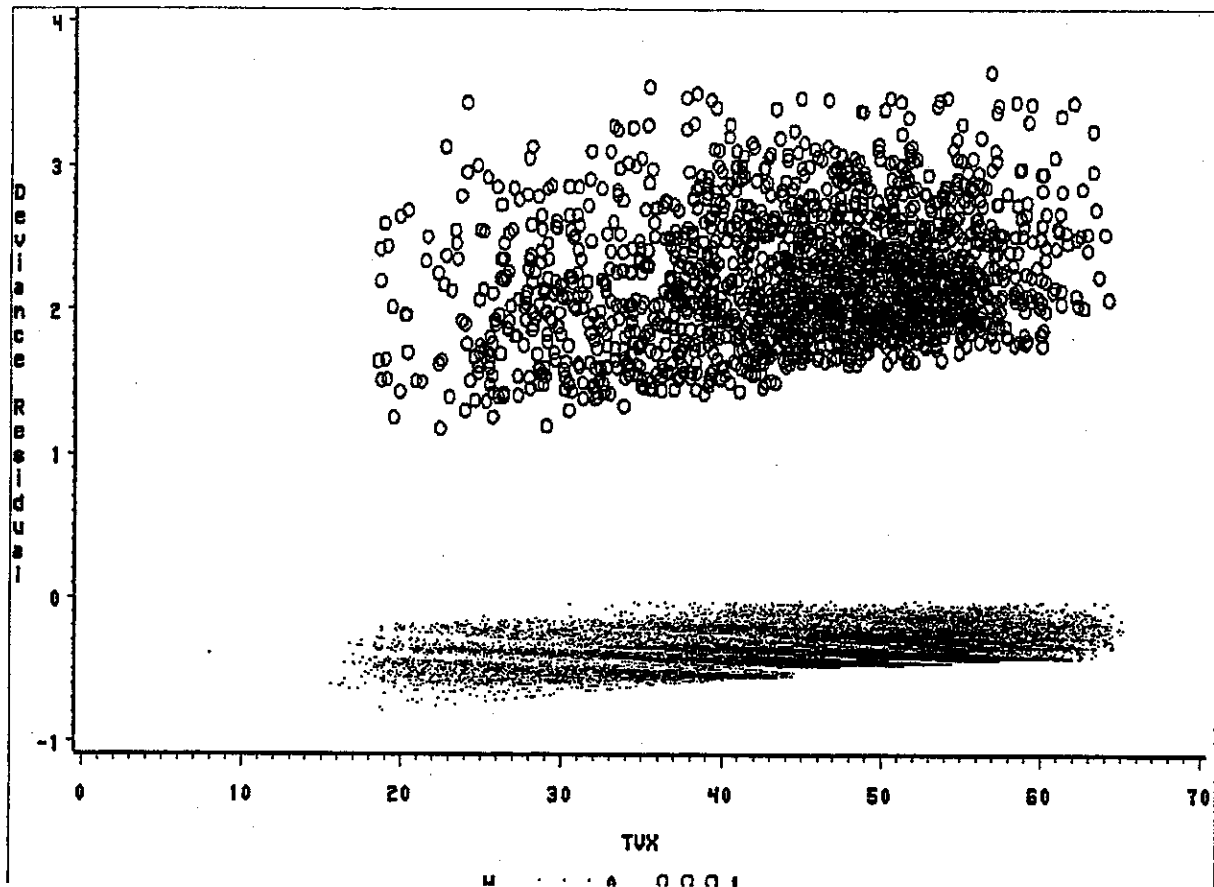
Kuvio T3.6 Työkyvyttömyyseläkkeeltä kuolleet, Weibull mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



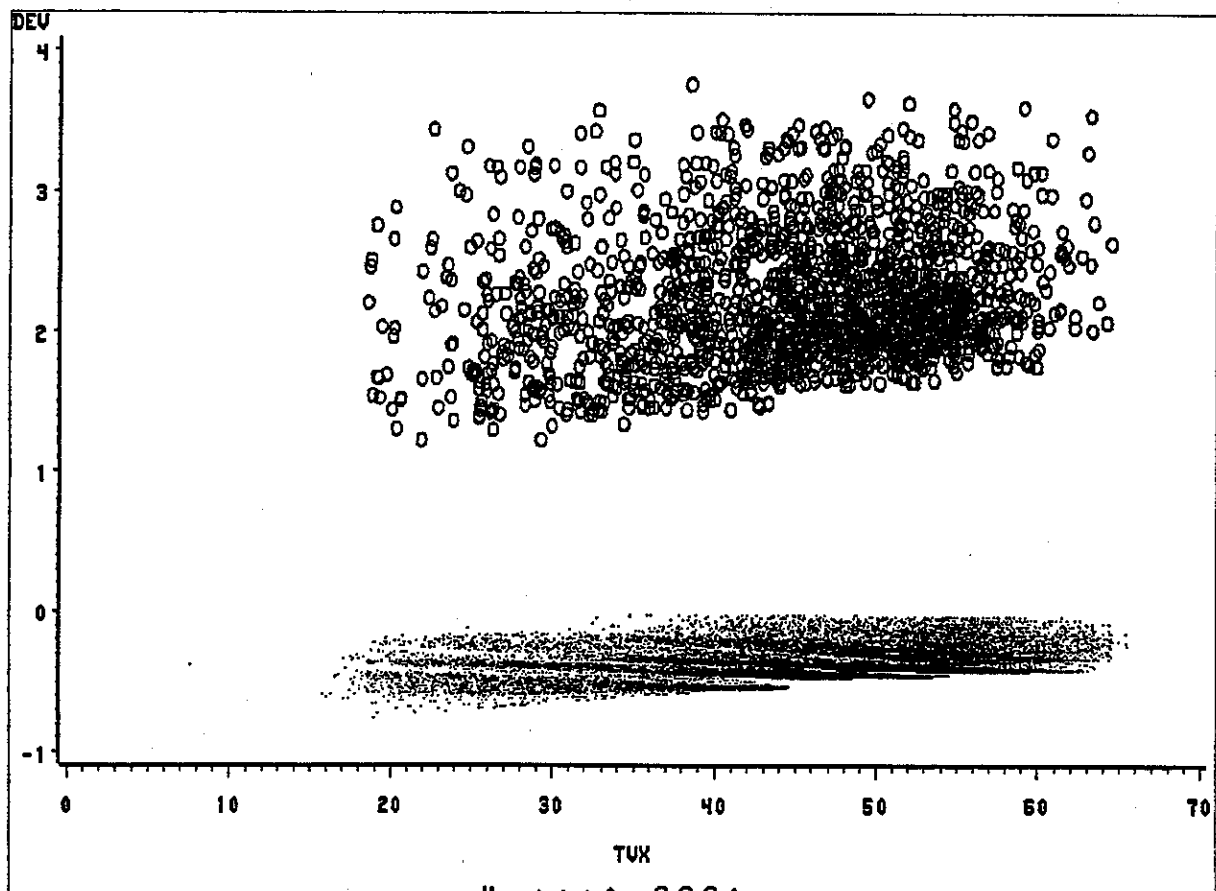
Kuvio T3.7 Työkyvyttömyyseläkkeeltä kuolleet, log-logistisen mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



Kuvio T3.8 Työkyvyttömyyseläkkeeltä poistuneet, Coxin mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)



Kuvio T3.9 Työkyvyttömyyseläkkeeltä poistuneet, log-logistisen mallin deviance-residuaalit (DEV) vastaan ikä tapahtumavuonna (TVX). Parantuneet merkitty ympyrällä (O), katotapaukset pisteellä (.)

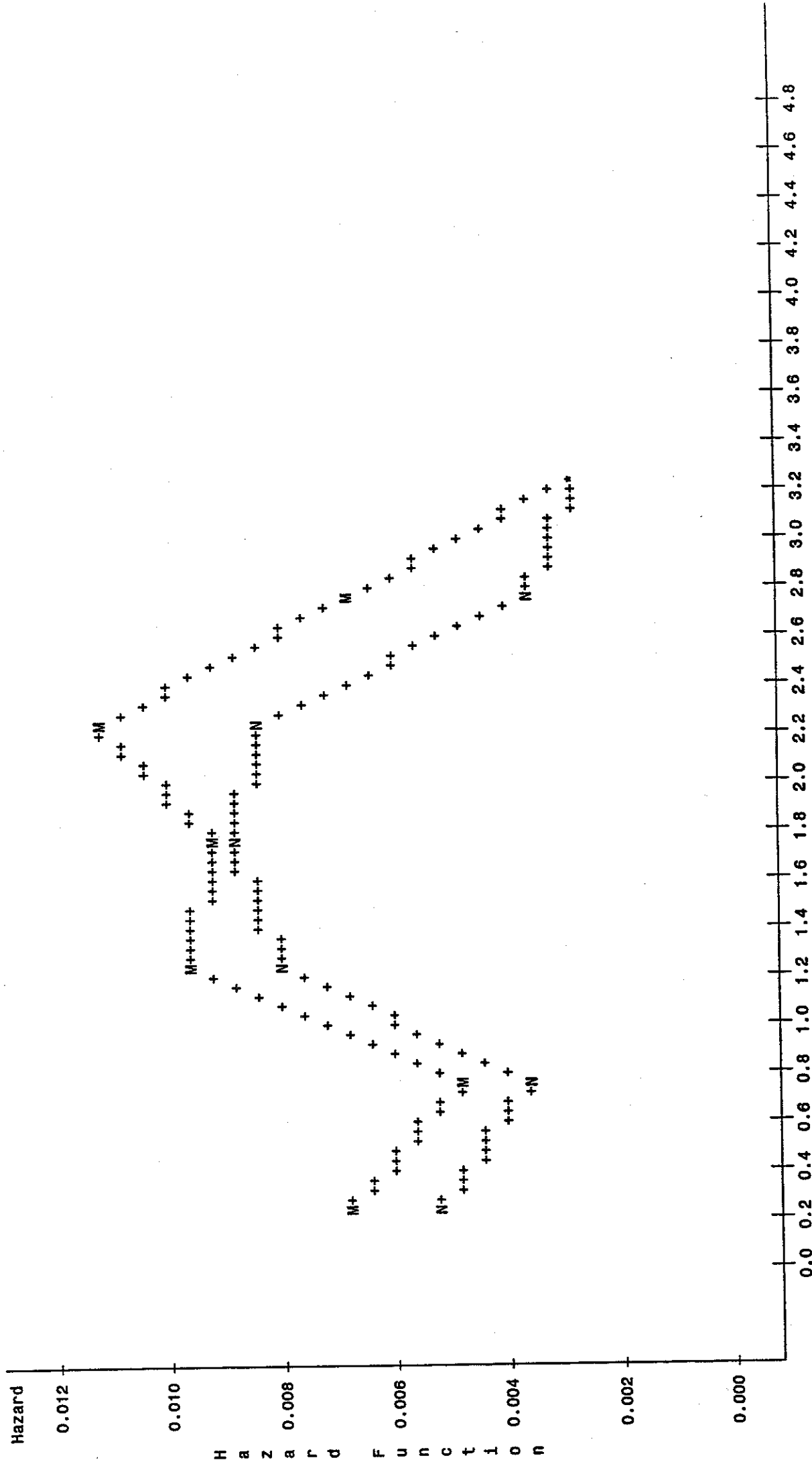


hite s kuns73.1

Parantuneet

The LIFETEST Procedure

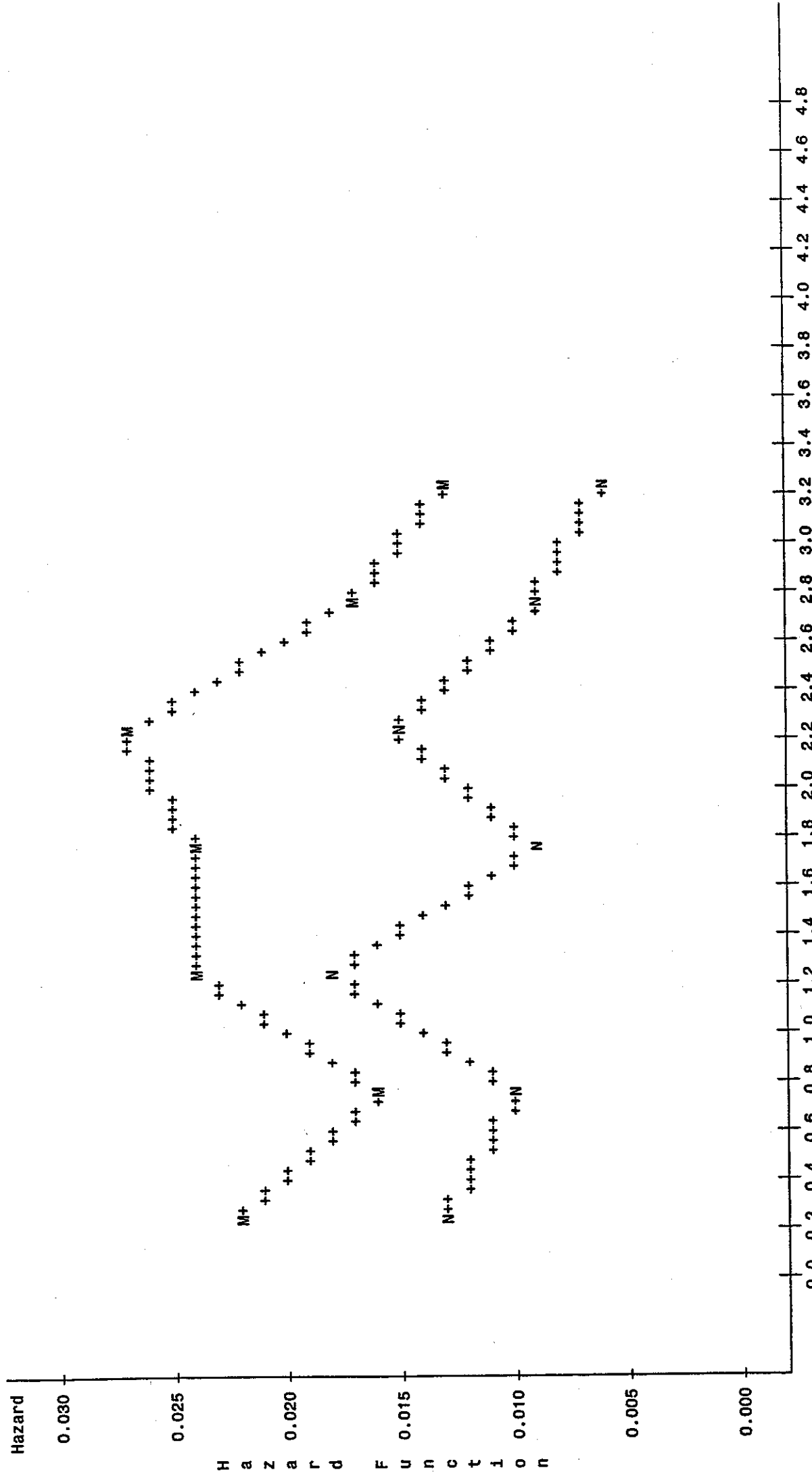
Hazard Function Estimates



Quollect

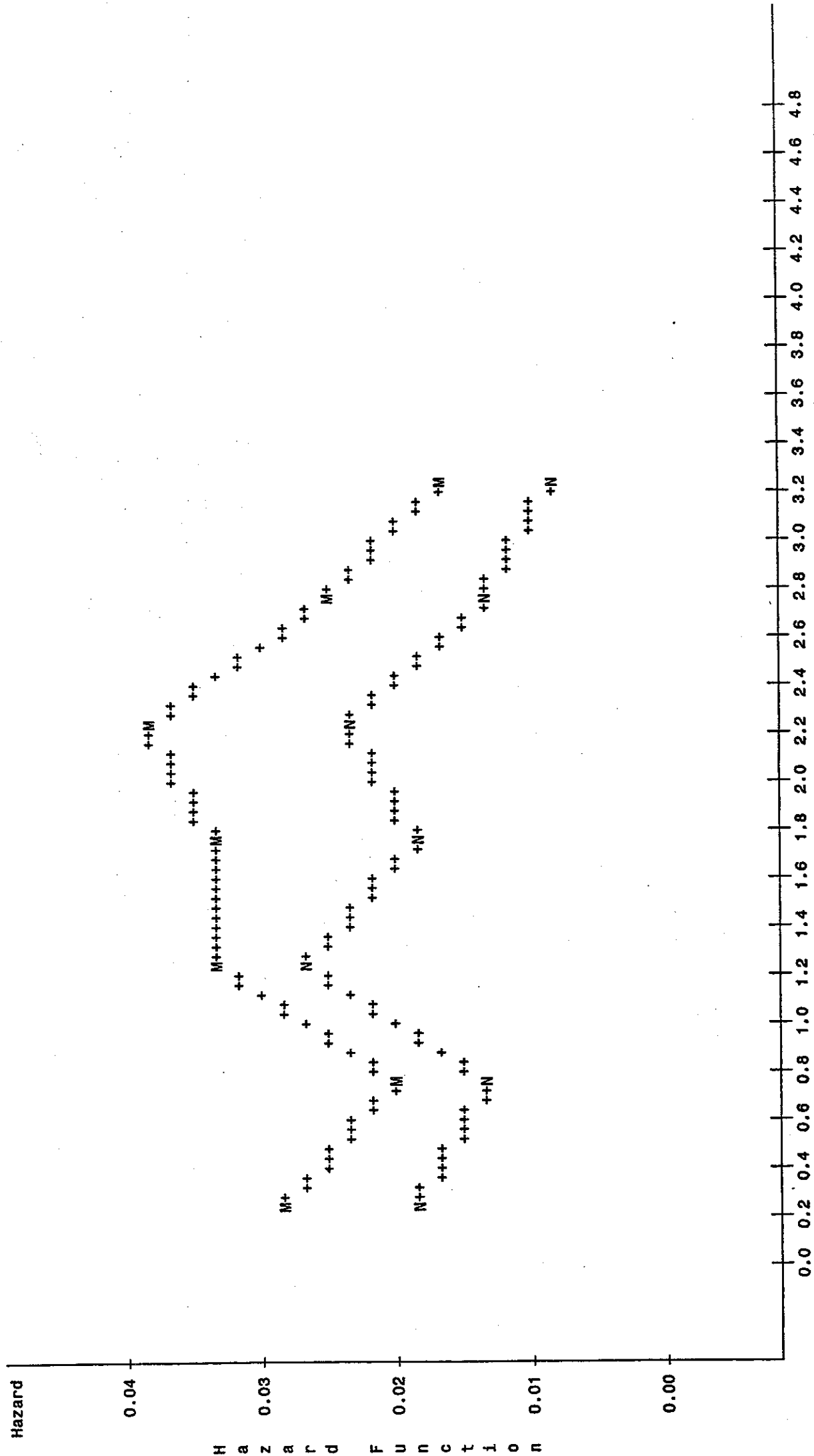
The LIFETEST Procedure

Hazard Function Estimates



The LIFETEST Procedure

Hazard Function Estimates



The PHREG Procedure

Data Set: WORK.DL4
 Dependent Variable: DD
 Censoring Variable: S
 Censoring Value(s): 0
 Ties Handling: BRESLOW

Summary of the Number of
 Event and Censored Values

	Total	Event	Censored	Percent Censored
	27828	540	27288	98.06

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	10897.075	9481.626	1415.449 with 5 DF (p=0.0001)
Score	.	.	1454.222 with 5 DF (p=0.0001)
Wald	.	.	769.210 with 5 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
SVL1940	1	1.397789	0.18214	58.89728	0.0001	4.046
SVL1950	1	1.458519	0.20480	50.72012	0.0001	4.300
SVL1960	1	0.970722	0.26613	13.30462	0.0003	2.640
SVJ	1	753.998198	37.20289	410.75918	0.0001	.
TVJ	1	3.851324	0.22891	283.06404	0.0001	47.055

Lifereg Procedure

Data Set =WORK.DL4
Dependent Variable=Log(DD)
Censoring Variable=S
Censoring Value(s)= 0
Noncensored Values= 534 Right Censored Values= 27237
Left Censored Values= 0 Interval Censored Values= 0
Observations with Zero or Negative Response= 57

Log Likelihood for LOGISTIC -2328.435332

The SAS System

Lifereg Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	16.9638038	0.897685	357.1073	0.0001	Intercept
SVL1940	1	-1.0599124	0.148715	50.79603	0.0001	
SVL1950	1	-1.0895898	0.168653	41.73857	0.0001	
SVL1960	1	-0.7477251	0.21574	12.01216	0.0005	
SVJ	1	-591.85431	36.22764	266.9005	0.0001	
TVJ	1	-3.0182283	0.205697	215.3012	0.0001	
SCALE	1	0.75085524	0.031098			Logistic scale parameter

Lifereg Procedure

Data Set =WORK.DL4
 Dependent Variable=Log(DD)
 Censoring Variable=S
 Censoring Value(s)= 0
 Noncensored Values= 534 Right Censored Values= 27237
 Left Censored Values= 0 Interval Censored Values= 0
 Observations with Zero or Negative Response= 57

Log Likelihood for WEIBULL -2326.364307

Lifereg Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	17.0280857	0.8973	360.1273	0.0001	Intercept
SVL1940	1	-1.0906164	0.148993	53.58146	0.0001	
SVL1950	1	-1.1203086	0.165325	45.91943	0.0001	
SVL1960	1	-0.7584888	0.207799	13.32334	0.0003	
SVJ	1	-588.31783	35.78457	270.2915	0.0001	
TVJ	1	-3.0145341	0.20479	216.6821	0.0001	
SCALE	1	0.76868742	0.032005			Extreme value scale parameter

The PHREG Procedure

Data Set: WORK.DL4
 Dependent Variable: DD
 Censoring Variable: Q
 Censoring Value(s): 0
 Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
27828	1327	26501	95.23

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates		With Covariates		Model Chi-Square
	Score	Wald	Score	Wald	
-2 LOG L	26792.157	26652.187	139.970 with 6 DF	(p=0.0001)	
Score	.	.	144.610 with 6 DF	(p=0.0001)	
Wald	.	.	139.887 with 6 DF	(p=0.0001)	

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
SPN	1	-1.675057	0.39739	17.76701	0.0001	0.187
TVL4	1	-0.327071	0.15867	4.24884	0.0393	0.721
TVJ	1	-0.216797	0.06526	11.03451	0.0009	0.805
SVJ	1	-70.544910	11.89166	35.19219	0.0001	0.000
TVC	1	0.270181	0.11976	5.09005	0.0241	1.310
SVC	1	80.369918	21.32556	14.20320	0.0002	8.021E34

The SAS System

Lifereg Procedure

Data Set =WORK.DL4
 Dependent Variable=Log(DD)
 Censoring Variable=Q
 Censoring Value(s)= 0
 Noncensored Values= 1307 Right Censored Values= 26464
 Left Censored Values= 0 Interval Censored Values= 0
 Observations with Zero or Negative Response= 57

Log Likelihood for LLOGISTIC -6261.875015

The SAS System

Lifereg Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	2.27681346	0.252453	81.2685	0.0001	Intercept
SPN	1	1.44767998	0.346924	17.41316	0.0001	
TVL4	1	0.27817006	0.140989	3.892671	0.0485	
TVJ	1	0.19118976	0.057767	10.9541	0.0009	
SVJ	1	60.4727987	10.51195	33.09434	0.0001	
TVC	1	-0.2412355	0.104141	5.365859	0.0205	
SVC	1	-67.323491	18.60779	13.09011	0.0003	
SCALE	1	0.84434138	0.022377			Logistic scale parameter

Lifereg Procedure

Data Set =WORK.DL4
 Dependent Variable=Log(DD)
 Censoring Variable=Q
 Censoring Value(s)= 0
 Noncensored Values= 1307 Right Censored Values= 26464
 Left Censored Values= 0 Interval Censored Values= 0
 Observations with Zero or Negative Response= 57

Log Likelihood for WEIBULL -6261.910887

The SAS System

09:01 Wednesday, January 7, 1998 16

Lifereg Procedure

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	2.35475366	0.248068	90.1054	0.0001	Intercept
SPN	1	1.42683531	0.344801	17.12424	0.0001	
TVL4	1	0.27444437	0.138646	3.918275	0.0478	
TVJ	1	0.18705159	0.056597	10.92295	0.0009	
SVJ	1	59.6268869	10.3698	33.06312	0.0001	
TVC	1	-0.2366238	0.103431	5.23377	0.0222	
SVC	1	-66.350331	18.52252	12.83176	0.0003	
SCALE	1	0.85635413	0.022833			Extreme value scale parameter

The PHREG Procedure

Data Set: WORK.DL4

Dependent Variable: DD

Censoring Variable: W

Censoring Value(s): 0

Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
27828	1867	25961	93.29

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	37689.231	37322.619	366.612 with 7 DF (p=0.0001)
Score	.	.	355.976 with 7 DF (p=0.0001)
Wald	.	.	334.756 with 7 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
SPN	1	-2.144545	0.40098	28.60377	0.0001	0.117
SVL1930	1	-0.352897	0.12595	7.85090	0.0051	0.703
SVL1940	1	-0.264153	0.08851	8.90633	0.0028	0.768
TVJ	1	0.260923	0.06258	17.39288	0.0001	1.298
SVJ	1	58.437265	13.28536	19.34786	0.0001	2.393E25
TVC	1	0.412730	0.11667	12.51370	0.0004	1.511
SVC	1	104.594772	18.29243	32.69469	0.0001	2.66E45

Lifereg Procedure

Data Set =WORK.DL4
 Dependent Variable=Log(DD)
 Censoring Variable=W
 Censoring Value(s)= 0
 Noncensored Values= 1841 Right Censored Values= 25930
 Left Censored Values= 0 Interval Censored Values= 0
 Observations with Zero or Negative Response= 57

Log Likelihood for LLOGISTIC -8077.272294

The SAS System

09:01 Wednesday, January 7, 1998 1

Lifereg Procedure

Variable DF Estimate Std Err ChiSquare Pr>Chi Label/Value

INTERCPT	1	3.64826122	0.225378	262.0281	0.0001	Intercept
SPN	1	1.80600743	0.344554	27.47418	0.0001	
SVL1930	1	0.32044139	0.111616	8.242218	0.0041	
SVL1940	1	0.23412199	0.078117	8.982471	0.0027	
TVJ	1	-0.2167684	0.054375	15.89257	0.0001	
SVJ	1	-48.746228	11.82351	16.99767	0.0001	
TVC	1	-0.3512752	0.100141	12.30459	0.0005	
SVC	1	-87.386455	15.85014	30.3984	0.0001	
SCALE	1	0.83092784	0.018623			Logistic scale parameter

Lifetest Procedure

Data Set =WORK.DL4
 Dependent Variable=Log(DD)
 Censoring Variable=W
 Censoring Value(s)= 0
 Noncensored Values= 1841 Right Censored Values= 25930
 Left Censored Values= 0 Interval Censored Values= 0
 Observations with Zero or Negative Responses= 57

Log Likelihood for WEIBULL -8074.119763

The SAS System

Lifetest Procedure

Variable DF Estimate Std Err ChiSquare Pr>Chi Label/Value

INTERCPT	1	3.77388042	0.223861	284.1972	0.0001	Intercept
SPN	1	1.82205997	0.345881	27.75058	0.0001	
SVL1930	1	0.31019576	0.109771	7.98534	0.0047	
SVL1940	1	0.22641337	0.07675	8.702626	0.0032	
TVJ	1	-0.2270316	0.053839	17.78209	0.0001	
SVJ	1	-50.407709	11.67012	18.65706	0.0001	
TVC	1	-0.3561723	0.100213	12.63208	0.0004	
SVC	1	-88.147801	15.80168	31.11833	0.0001	
SCALE	1	0.84943765	0.019208			Extreme value scale parameter