# Henkivakuutusyhtiöiden kuolevuustutkimus K2012

17.12.2012


- Julkaisija
  Finanssivalvonta



- Tekijä
  Tarja Sirén, 010 831 5515



- Julkaisun nimi
  Henkivakuutusyhtiöiden kuolevuustutkimus K2012



- Tiivistelmä – Referat – Summary
  Tämä raportti sisältää pitkän ajan kuolevuusennusteen Suomessa vapaaehtoisesti henkivakuutetuille vakuutuslajeittain. Ennuste on tehty kahdessa vaiheessa. Ensiksi tehtiin pitkän ajan kuolevuusennuste Lee-Miller-menetelmällä Suomen väestölle sukupuolittain. Toiseksi estimoitiin korjauskertoimet vakuutuslajeittain Poisson-regressiolla. Näiden tulosten perusteella muodostettiin henkivakuutuksen vakuutuslajikohtainen pitkän ajan kuolevuusennuste sekä arvioitiin ennusteen luottamusvälit.

  Denna rapport innehåller en långsiktig dödlighetsprognos över personer med frivillig livförsäkring i Finland efter försäkringsgren. Prognosen har tagits fram i två steg. Först utarbetades en dödlighetsprognos enligt kön för Finlands befolkning med hjälp av Lee-Miller-metoden. Därpå estimerades korrigeringskoefficienter efter försäkringsgren genom Poissonregression. På basis av resultaten utformades en långsiktig dödlighetsprognos efter försäkringsgren inom livförsäkringssektorn och analyserades konfidensintervallen för prognosen.

  This paper contains long term mortality forecast for voluntary life insurance by product lines in the Finnish market. The mortality forecast for life insurance has been computed in two steps. First, we calculated long term mortality forecast with Lee-Miller method for Finnish population data. Second, we estimated correction terms for each life insurance product line via Poisson-regression. Based on these results we formed long term mortality forecast and confidence intervals for voluntary life insurance product lines.



- Avainsanat – Nyckelord – Key words
  Vapaaehtoinen henkivakuutus, kuolevuusennuste, Lee-Miller, Poisson-regressio
  Frivillig livförsäkring, dödlighetsprognos, Lee-Miller, Poissonregression
  Voluntary life insurance, mortality forecast, Lee-Miller, Poisson-regression

- Sarja/nimi ja numero
  Tutkimukset 2012. ISSN 1798-1654. ISBN 978-952-5841-11-4

- Sivumäärä
  86

# Abstract

Life insurance companies need mortality forecast as they face mortality risk in their life insurance and longevity risk in their pension insurance businesses. This paper contains long term mortality forecast for voluntary life insurance by product lines in the Finnish market. The report discusses the background of mortality modeling and gives a review of some important references and viewpoints. In addition, it provides a general overview of voluntary life insurance.

The mortality forecast for life insurance is computed in two steps, because of the data limitations. First, we estimate long term mortality forecast with Lee-Miller method for Finnish population data. Second, we calculate correction terms for each life insurance product line via Poisson-regression. Based on these results we calculate long term mortality forecast and confidence intervals for voluntary life insurance product lines.

Correction terms used in the calculation of mortality forecast can be computed from count or volume data. In this report we concentrate on count data because of, among other things, data limitations. Nevertheless we still compare the correction terms based on both approaches. In case of males' group pension insurance we present results of count and volume data because there was a clear difference between the two.

Previous research with this focus was made by Mika Mäkinen in 2004. The reference mortality based on that research was published in February 2005. Our paper employs more recent and extensive data and takes into account the inherent heterogeneity in life insurance data by analyzing each product line separately. Hence we are able to provide the mortality forecast individually for each product line.

# Sisältö

**Abstract** **1**

**1 Johdanto** **5**

**2 Tausta** **7**
2.1 Kirjallisuuskatsaus . . . . . 7
2.2 Henkivakuutuksen referenssikuolevuus K2004 . . . . . 11
2.3 Lakisääteisen tapaturmavakuutuksen ja liikennevakuutuksen kuolevuusreferenssimalli . . . . . 12
2.4 Solvenssi I ja Solvenssi II . . . . . 13

**3 Henkivakuutusaineisto** **14**
3.1 Riskin valinta . . . . . 14
3.2 Tutkimusaineisto . . . . . 14
3.3 Kohortin koon vaikutus . . . . . 16
3.4 Vastuunvalinnan vaikutus . . . . . 16
3.5 Säästöhenkivakuutukset . . . . . 17
3.6 Riskivakuutukset . . . . . 18
3.7 Eläkevakuutukset . . . . . 19
3.8 Perhe-eläke . . . . . 20
3.9 Tapaturmainen kuolema . . . . . 20
3.10 Yhteenveto . . . . . 21

**4 Väestökuolevuuden ennustaminen** **23**
4.1 Yleistä . . . . . 23
4.2 Menetelmän valinta . . . . . 23
4.3 Havaintoaineisto . . . . . 26
4.3.1 Lähtöaineiston valinta . . . . . 26
4.3.2 Lähtövuosien valinta . . . . . 26
4.4 Ennusteen epävarmuus . . . . . 31
4.5 Lopullinen ennuste ja vertailut aikaisempiin tutkimuksiin . . . 32

**5 Henkivakuutuksen korjauskertoimet** **36**
5.1 Yleistä . . . . . 36
5.2 Lukumäärä- ja summakuolevuuksien vertailu . . . . . 36
5.3 Poisson-jakauma . . . . . 39
5.4 Poisson-regressio . . . . . 39
5.5 Korjauskertoimen määrittäminen . . . . . 41
5.6 Soveltaminen henkivakuutusaineistoon . . . . . 44
5.6.1 Yleistä . . . . . 44
5.6.2 Henkivakuutusaineisto . . . . . 47

5.6.3 Yksilölliset säästövakuutukset . . . . . 48
5.6.4 Yksilöllinen riskivakuutus . . . . . 49
5.6.5 Ryhmäetuvakuutus ja ryhmäriskivakuutus . . . . . 50
5.6.6 Yksilöllinen eläkevakuutus ja perhe-eläke . . . . . 51
5.6.7 Ryhmäeläkevakuutus . . . . . 52
5.6.8 Tapaturmainen kuolema . . . . . 53
5.7 Yhteenveto . . . . . 54

**6 Raportin yhteenveto 61**

**Lähteet 66**

**A Aineiston kuvaus 69**

**B Väestökuolevuusennustetta vastaavat elinajanodotteet eri lähtövuosilla 71**

**C Vertailut muiden tutkimuksien kanssa 72**

**D $m_{x,t}$-, $p_{x,t}$- ja $e_x$-luvut 74**

**E Henkivakuutuskuolevuuden epävarmuuden huomioon ottaminen 75**

**F Summakuolevuuksien sovitteet 76**

**G K2004-tutkimusta vastaavan henkivakuutuskuolevuuden ja sovitteen suhteellinen ero 79**

**H R-ohjelmiston ohjeet 81**
H.1 Tavoite . . . . . 81
H.2 R-ohjelman alustus . . . . . 82
H.3 Kuolevuus . . . . . 83
H.3.1 Aineiston lataaminen ja valinta . . . . . 83
H.3.2 Kuolevuusennuste . . . . . 84
H.3.3 Elinajanodotteiden laskenta . . . . . 85
H.4 Poisson-regressio . . . . . 85

# 1 Johdanto

Syksyllä 2011 käynnistettiin Fivan valvonta- ja tutkimustoiminnan tarpeisiin tutkimusprojekti (K2012) analysoimaan henkivakuutuksen nykyistä referenssikuolevuutta K2004 uuden datan ja uusien mallien avulla ottaen myös huomioon muut vakuutusalalla viime aikoina tehdyt kuolevuustutkimukset. Hankkeella on liittymäkohtia muun muassa Solvenssi II -projektiin ja pohjoismaiseen yhteistyöhön.

Projektin tavoitteena oli

1. kerätä henkivakuutusyhtiöiltä vanhan riskiperusteanalyysin mukainen kuolevuusaineisto 2002–2010 vakuutustasolla. Vuosien 1972–2001 summatason karkeampi aineisto oli käytettävissä K2004 -projektista.

2. analysoida aineisto stokastisen mallin avulla

3. laatia kuolevuuden pitkän aikavälin ennuste ottaen huomioon koko väestöä koskeva tilastoaineisto (Human Mortality Database).

4. laatia simulointialgoritmi

5. selvittää tarve ja mahdollisuus referenssiperustekuolevuudelle sekä laatia tarvittavat ehdotukset, jotka voivat toimia sekä tarkastuksen että henkivakuutusyhtiöiden riskienhallinnan ja aktuaaritoimintojen apuvälineinä.

6. laatia yksityiskohtainen tutkimusraportti

7. luoda pohjaa mahdollisille Fivan jatkotutkimuksille (esim. muut mallit)

8. edistää yhteistyötä henkivakuutusalan kanssa.

Projektiorganisaation ohjausryhmän muodostivat Vesa Hänninen, Jari Niittuinperä, Hely Salomaa ja Vesa Ronkainen Finanssivalvonnasta. Projektiryhmän muodostivat Vesa Ronkainen (projektin vetäjä), Tarja Sirén (tutkimuksen toteuttanut tutkija) sekä henkivakuutusaktuaareina Jari Niittuinperä ja Mika Mäkinen. Tukiryhmän muodostivat henkivakuutusyhtiöiden vastuulliset aktuaarit tai heidän edustajansa. Akateemisena asiantuntijana oli Arto Luoma Tampereen Yliopistolta.

Raportti alkaa motivoinnilla, jossa käydään läpi kuolevuustutkimuksen taustaa ja tarkoitusta sekä annetaan lyhyt katsaus tärkeimpiin lähteisiin ja näkökulmiin (luku 2). Luvussa 3 analysoidaan henkivakuutusyhtiöiltä kerättyä vakuutusaineistoa. Koska saatavilla oleva henkivakuutusaineisto ei ole riittävä pitkän aikavälin ennusteen tekemiseen, tehdään kuolevuusennuste

Suomen väestöaineistolla Lee-Miller -menetelmällä, jota korjataan henkivakuutusaineistosta lasketulla korjauskertoimella. Luvussa 4 perustellaan ja analysoidaan Suomen väestöaineistolla tehty kuolevuusennuste. Aineisto on saatavilla HMD-tietokannasta osoitteesta `www.mortality.org`. HMD-tietokantaan on kerätty populaatio- ja kuolevuusaineistoa 37 maalta. Suomen aineisto perustuu Tilastokeskuksen tilastoihin. Luvussa 5 määritellään korjauskerroin, jonka avulla Suomen väestön kuolevuusennustetta korjataan henkivakuutettujen kuolevuutta vastaavaksi. Henkivakuutustuotteiden heterogeenisuuden vuoksi analyysit tehdään tuoteryhmittäin. Lisäksi luvussa tuodaan esille korjauskertoimeen liittyviä haasteita, kuten aineistoon liittyvät rajoitukset. Lopuksi luvussa 6 esitetään tulosten yhdistäminen sekä vedetään yhteen raportin tärkeimmät päätelmät. Liitteessä C verrataan väestökuolevuusennusteen tuloksia muihin tutkimuksiin. Liitteessä D on $m_{x,t}$-taulukot hajontoineen. Liitteessä E esitetään, miten henkivakuutuskuolevuuden epävarmuus voidaan ottaa huomioon.

Laskelmat on tehty pääosin R-ohjelmistolla (versio 2.15.0). Kuolevuusennusteet on tehty R-ohjelmiston Rob J. Hyndmanin demography-paketin versiolla 1.13 (ladattavissa osoitteesta `http://robjhyndman.com/software/demography`). Liitteessä H on ohjeita ja esimerkkejä R-ohjelmiston käytöstä.

# 2 Tausta

## 2.1 Kirjallisuuskatsaus

Kuolevuuden laskeva trendi, joka näkyy kuvassa 2.2, on havaittu maailmanlaajuisesti kehittyneissä maissa. Tällä kehityksellä on suuria yhteiskunnallisia ja taloudellisia vaikutuksia (Holzmann & Palmer (2006)). Haasteita kohtaavat muun muassa terveydenhuolto- ja eläkevakuutusjärjestelmät. Toisaalta haasteena on myös tulevien kustannusten realistinen arvioiminen ja niihin varautuminen.

Tästä syystä pitkäikäisyyden/kuolevuuden mallintaminen on tärkeää. Useat tutkimukset eri maissa osoittavat, etteivät väestötieteilijöiden ja aktuaarien yleisesti käyttämät perinteiset deterministiset menetelmät ole ottaneet huomioon riittävästi kuolevuuden alenemista ja ennusteisiin liittyvää epävarmuutta (esim. Keilman (1990), Alho (1990), Keilman et al. (2008)). Lisäksi on havaittu, että asiantuntija-arviot eivät ole tuoneet ennusteisiin lisäarvoa puhtaisiin tilastollisiin malleihin verrattuna vaan asia on ollut päinvastoin (Alho et al. (2012)).

Kuolevuusmallinnus on kehittynyt huomattavasti viime vuosina ja painottunut erityisesti stokastisiin malleihin. Erilaisia mallivaihtoehtoja on paljon, ja ne jakaantuvat eri käyttötarkoitusten mukaan (Booth & Tickle (2008), Cairns et al. (2008), Keilman et al. (2008)). Mallinvalinnassa päätekijöitä ovat muun muassa

- tarkasteltava populaatio (koko väestö tai vakuutuskanta vai sen osa)
- stokastinen vai deterministinen lähestymistapa
- selittävät muuttujat (esim. sukupuoli, ikä, vastuunvalinta ja kuolinsyy)
- ennustejakson pituus
- datan valinta ja saatavuus

Tyypillisimmät kuolevuusmallit voidaan esittää muodossa $f(x_1)$, $f(x_1, x_2)$ ja $f(x_1, x_2, x_3)$, missä $x_1$=ikä, $x_2$=aika, $x_3$=syntymäkohortti. Esimerkiksi Lee-Carter -malli on kahden faktorin eli bilineaarinen malli (Booth & Tickle (2008)).

Kuolevuusriskin kehityksen taustalla on siis monia eri tekijöitä kuten syntymäaika, sukupuoli, elintavat, sosioekonominen luokka, geneettinen perimä, sairaudet ja niin edelleen. Näiden tekijöiden vaikutusta kuolevuuteen on tutkittu paljon (ONS (2009)). Kuitenkin on haasteellista tutkia kunkin tekijän yksittäistä vaikutusta kuolevuuteen ja ymmärtää niiden yhteisvaikutus.

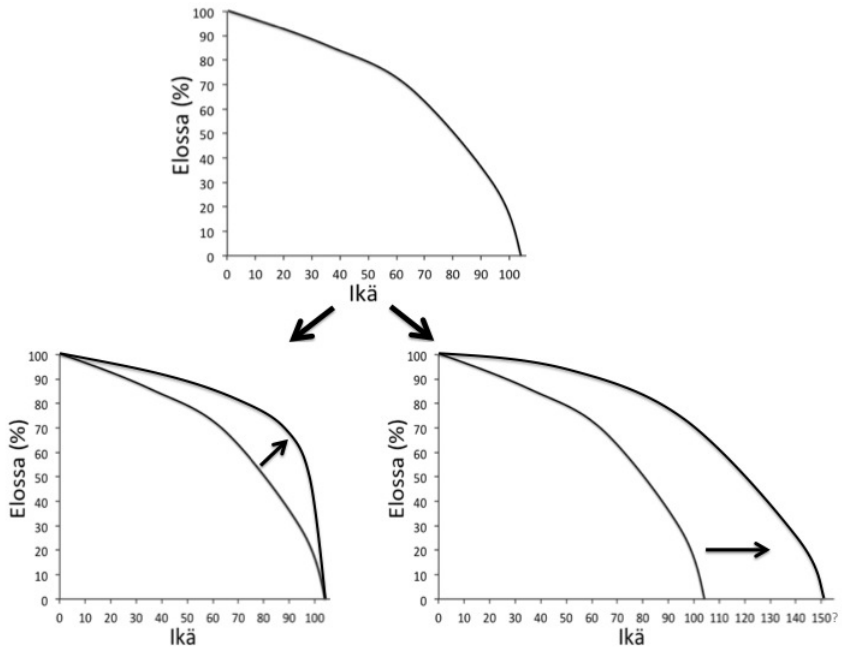
Lisäksi riittävän yksityiskohtaisen ja laajan aineiston saatavuus asettaa rajoitteita analyysille. Toisaalta on vaikea ennustaa, mitkä tulevat olemaan merkittävimmät kuolinsyyt tulevaisuudessa, miten elintapojen muutokset tulevat vaikuttamaan kuolevuuteen tai mikä on mahdollisten lääketieteellisten läpimurtojen vaikutus kuolevuuden kehitykseen. Näin ollen esimerkiksi kuolinsyiden muutosten mallintaminen pitkän ajan ennusteissa on haasteellista. Kuitenkin joidenkin tuotteiden hinnoittelussa ja lyhyen ajan ennusteissa kuolinsyiden huomioon ottaminen voi olla hyvinkin perusteltua (Alho & Spencer (2005), Girosi & King (2008)). Lisäksi merkittävien lääketieteellisten havaintojen vaikutusten analysointi olisi järkevää (Yrjö Turtiainen SHY (1982)). Tällaisia voisivat olla esimerkiksi lisääntyneen vanhuusiän diabeteksen tai eräiden syöpähoitojen kehittymisen vaikutus kuolevuuteen. Tutkimustyön tulisikin olla poikkitieteellistä.

Edellä mainittuja ominaisuuksia huomioon ottavia malleja kutsutaan selittäviksi malleiksi, jotka perustuvat lääketieteen syy-seuraus -suhteen malleihin eri riskitekijöille ja kuolinsyille. Tilastolliset mallit perustuvat oletukseen, että tulevaisuuden kehitystä voidaan ekstrapoloida historiallisen datan ja tilastollisten mallien avulla (Booth & Tickle (2008)).

Asiantuntijoilla on aina ollut eriäviä mielipiteitä siitä, miten kuolevuuden kehitys jatkuu tulevaisuudessa (Alho et al. (2012)). Kehittyneissä maissa kuolevuuden aleneminen alkoi imeväisyyskuolevuuden alentumisella, minkä jälkeen alenivat aikuisiän kuolevuus ja viimeisimpänä iäkkäiden kuolevuus. Eliniän pitenemiseen ovat vaikuttaneet voimakkaasti puhtaan veden saanti, alueellinen ja sukupuoleen liittyvä genetiikka sekä elintavat. Kehittyneissä maissa miesten ja naisten elintavat yhdenmukaistuvat, jolloin oletettavasti kuolevuusero kapenee. Eliniän ennustamiseen pitkällä aikavälillä vaikuttaa olennaisesti lääketieteen kehitys. Tulevaisuuden kehityssuuntia ovat muun muassa kudossiirteet, kantasolutekniikat, geenimanipulaatiot jne. Toisaalta saasteiden lisääntyminen, liika lihavuus kehittyneissä maissa ja elintapasairaudet saattavat vaikuttaa eliniän kehitykseen negatiivisesti.

Kuvassa 2.1 on esitetty kaksi mahdollista tulevaisuuden skenaariota. Vasemmanpuoleinen on niin sanottu ”squaring-off” -skenaario. Skenaariossa maksimaalinen elinikä pysyy vakiona mutta muiden ikäluokkien elinajanodote kasvaa. Oikeanpuoleinen on niin sanottu ”expansion”-skenaario, missä myös maksimaalinen elinikä kasvaa. Jälkimmäisen todennäköisyyteen vaikuttaa muun muassa tulevaisuudessa mahdollisesti tapahtuvat lääketieteelliset läpimurrot.

Kysymys, onko ihmisen eliniällä biologista maksimia, ei ole selvä. Vajaat 200 vuotta sitten tehtyjen arvioiden mukaan maksimaalinen elinikä olisi noin 125 vuotta (Flourens (1850)). Tämä on melko lähellä nykyistä havaittua maksimimaalista elinikää (tällä hetkellä vanhimmaksi elänyt henkilö oli yli 122-vuotias). Toisaalta viime vuosikymmeninä tutkimus telomeerien lyhentymisen vaikutuksesta vanhenemiseen ja syöpään on kehittynyt huomattavasti. Myös telomeerien lyhentymisen ehkäiseviä menetelmiä kehitellään (Greider & Blackburn (1985)). Näin ollen kysymys ihmisen eliniän biologisesta maksimista ei ole selvä, ja se jakaa tiedemiesten mielipiteitä. Lisäksi tulee muistaa, että elinajanodote jää vielä pitkän aikaa alle edellä mainitun maksimielinajan. Näin ollen elinajanodotteella on edelleen tilaa kasvaa. Viime vuosina



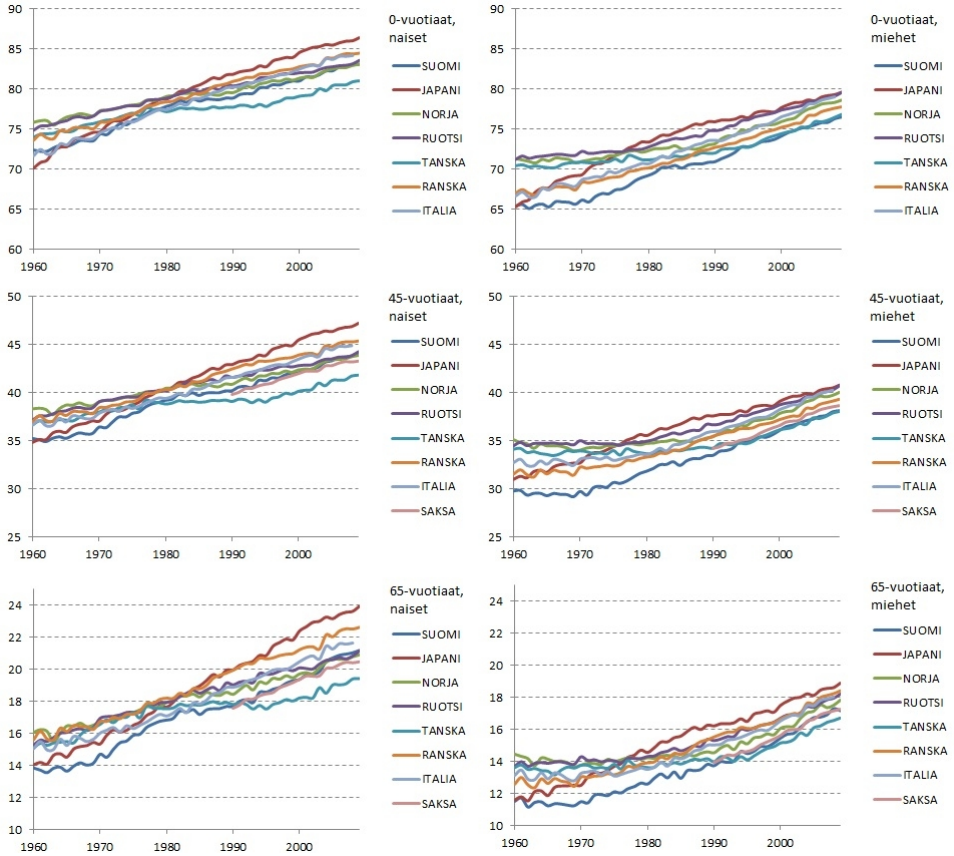

Kuva 2.1: Elinajanodotteen kehityksen eri skenaariot

Japani on johtanut elinajanodotteen kehityksessä niin miesten kuin naistenkin osalta. Oeppen & Vaupel (2002) esittivät niin sanotun ”best-practice” -suoran, jonka perusteella naisten suurin elinajanodote näyttäisi seuraavan suoraa maailmanlaajuisesti. Tutkimuksen perusteella ei ole viitteitä siitä, että elinajanodotteen kehityssuunta olisi muuttumassa lähivuosina, koska sen tulisi näkyä ensiksi ”best-practice” -maissa.

Vaikka kuolevuuden laskeva trendi on havaittu maailmanlaajuisesti kehittyneissä maissa, maiden välinen kehitys poikkeaa toisistaan paljon (Rau et al. (2006) ja Rau et al. (2008)). Toisin sanoen kuolevuuden alenemisen tahti vaihtelee maittain huomattavasti, varsinkin jos vertailuja tehdään sukupuolittain ja/tai ikäluokittain. Joidenkin tutkimusten pohjalta ilmiö näyttää olevan hetkellinen (Robine et al. (2005)). Tähän viittaa myös kuva 2.2, josta huomataan myös iän vaikutus maiden välisiin eroihin. Gaille & Sherris (2010) esittivät artikkelissaan, että viiden yleisimmän kuolinsyyn (verenkierto-, hengityselintaudit, syöpä, ulkoiset syyt, lois- ja infektiotaudit) trendi on sama monissa maissa, vaikka pitkäaikainen stokastinen trendi saattaa vaihdella maittain. Tutkimuksessa olivat mukana USA, Australia, Sveitsi, Japani, Singapore, Italia, Norja, Ruotsi ja UK.

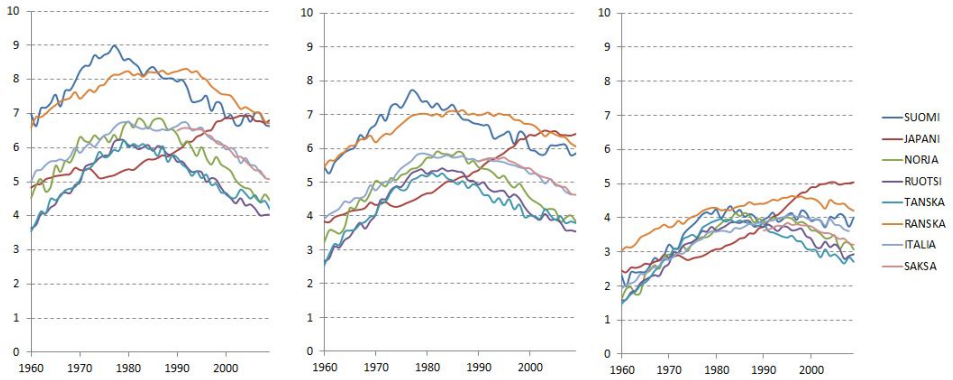
Toistaiseksi ei ole näyttöä, että maitten elinajanodotteet lähenisivät pysyvästi toisiaan esimerkiksi kehittyneissä maissa, vaikka tällainen hypoteesi voisi tuntua intuitiivisesti luontevalta. Tilanteissa, joissa populaatio on liian suppea pitkän ajan ennusteen tekemiseen, sopivasti valittujen maiden kuolevuusaineistojen yhdistäminen voisi olla mielekästä (Girosi & King (2008)). Sopivien maiden valinnassa on monenlaisia lähestymistapoja kuten kuolevuudeltaan mahdollisimman samanlaisten maiden valitseminen tai esimerkiksi tulevaisuuden mahdollisen kehityksen mielessä sopivien maiden yhdistäminen.

Kuvassa 2.3 on esitetty naisten ja miesten elinajanodotteiden erotus 0-,




Kuva 2.2: Elinajanodotteet maittain 0-, 45- ja 65-vuotiaille

45- ja 65-vuotiailla. OECD-maissa naisilla on kaikissa ikäluokissa korkeampi elinajanodote kuin miehillä. Syihin ei ole selvää vakiintunutta kantaa, mutta biologiset syyt ja tupakanpoltto nostetaan usein esiin. Japania lukuun ottamatta vastasyntyneiden ja 45-vuotiaiden elinajanodotteiden ero on suurin piirtein samalla tasolla kuin vuonna 1960 oltuaan välillä merkittävästi isompi. Kehitykseen voi olla useita syitä, mutta eräs lienee yhteiskunnan tasaarvoistuminen myös siinä mielessä, että miesten ja naisten elintavat ovat lähentyneet toisiaan. Maiden välisessä tarkastelussa Suomella, Japanilla ja Ranskalla elinajanodotteen ero miehillä ja naisilla on suurin.




Kuva 2.3: Elinajanodotteiden ero sukupuolittain 0-, 45- ja 65-vuotiaille

## 2.2 Henkivakuutuksen referenssikuolevuus K2004

Tällä hetkellä voimassa oleva henkivakuutuksen referenssikuolevuus K2004 perustuu Mika Mäkisen SHV-työhön (Mäkinen (2004)). Virallinen henkivakuutuksen referenssikuolevuus K2004 tehtiin Suomen Vakuutusyhtiöiden Keskusliiton (nykyisin Finanssialan Keskusliitto) toimesta aktuaariryhmässä, ja se hyväksyttiin 14.2.2005. Kuolevuusmallille K2004 oli seuraavia tavoitteita:

- kuvata hyvin kuolevuusilmiö
- kuolevuuden ikäkohtaisen muutoksen huomioon ottaminen
- sovellettavuus mahdollisimman moniin henkivakuutusluokkiin
- funktiomuotoinen esitystapa
- mallin tarkkuus erityisesti ikäluokissa 30–60 v.
- mallin käyttökelpoisuus

Referenssikuolevuudessa lähtöaineistona käytettiin henkivakuutusyhtiöiden riskiperusteanalyysejä ja aikaisempia kuolevuustutkimuksia vuosilta 1972–2001, jotka sisältävät yksilöllisen henkivakuutuksen havaitut kuolevuudet viisivuotisluokittain. Suomen väestöaineistona käytettiin HMD-tietokannan aineistoa vuosilta 1960–2001. Henkivakuutusaineiston suppeuden vuoksi referenssikuolevuus tehtiin ensin yleisesti tunnetulla Lee-Carter -menetelmällä Suomen väestökuolevuuden perusteella, minkä jälkeen henkivakuutusaineiston perusteella arvioitiin väestö- ja henkivakuutuskuolevuuden suhteellinen ero ja sen muutos ikäluokittain kaavalla:

$$p_x = \frac{\mu_x^{\text{henkiaineisto}}}{\mu_x^{\text{populaatio}}}.$$

Tutkimuksessa tehtiin oletus, että alle 20-vuotiailla ja yli 70-vuotiailla henkivakuutuskuolevuus vastaa populaation kuolevuutta. Malli tehtiin erikseen miehille ja naisille, kun taas K2004-referenssikuolevuutta edeltävässä K87-kuolevuudessa kuolevuusfunktio tehtiin miehille ja naisten kuolevuusfunktio saatiin määriteltyä ikäsiirtojen avulla.

Saatuihin kuolevuustauluihin ja ennusteisiin sovitettiin funktiosovite, jossa syötearvoina olivat ikä, syntymävuosi ja sukupuoli. Sovitteessa käytetyssä kuolevuusennusteessa ei otettu huomioon ennusteen epävarmuutta. Sovite tehtiin mahdollisimman tarkaksi ikävälille 30–60. Sovitteessa jätettiin ottamatta huomioon alle 20-vuotiaat. Malliin on sisällytetty korjaustermi, jonka avulla vanhempien ja nuorempien kuolevuus otetaan huomioon tarkemmin. Kuolevuudet esitettiin kohorteittain Makeham-funktion muodossa seuraavasti:

$$\mu_{sp}(x, sv) = \max\{0.0001; a_{sp}(sv)\} + e^{c_{sp}(x,sv)},$$

missä sukupuolesta riippuvat (sp=0 miehillä ja sp=1 naisilla) funktiot $a$ ja $c$ ovat

$$a_0(sv) = 10^{-5} \cdot 0.744 \cdot (2070 - sv)$$
$$a_1(sv) = 10^{-5} \cdot 0.206 \cdot (2019 - sv)$$

$$\begin{aligned} c_0(x, sv) &= 0.05438 \cdot (1716 - sv) + 0.000533 \cdot (sv - 1719) \cdot x - \\ &\quad 0.000217 \cdot (sv - 1843) \cdot \max\{0; x - 81\} \\ c_1(x, sv) &= -11,51 + 0.000316 \cdot (2253 - sv) \cdot x + \\ &\quad 0.000783 \cdot (sv - 1916) \cdot \max\{0; x - 71\} \end{aligned}$$

## 2.3 Lakisääteisen tapaturmavakuutuksen ja liikennevakuutuksen kuolevuusreferenssimalli

Suomalaisten vahinkovakuutusyhtiöiden uusi kuolevuusmalli julkaistiin vuonna 2011 (PricewaterhouseCoopers (2011)). Projekti toteutettiin kahdessa osassa. Ensimmäisessä vaiheessa tutkittiin Mira Kaupin tutkimukseen perustuvan

silloisen kuolevuusmallin päivitystarve (Kauppi (2004)). Toisessa vaiheessa tuotettiin uusi kohorttikohtainen kuolevuusreferenssimalli lakisääteistä vahinkovakuutusta harjoittavien vakuutusyhtiöiden käyttöön. Lisäksi toisessa vaiheessa julkaistiin ennusteen epävarmuuskertoimet sekä tuotettiin kuolevuuden korjauskertoimet, jotka toteutettiin Poisson-regressiolla. Raportissa julkaistiin myös mallin avulla lasketut elinajanodotteet ja eri korkotasoilla lasketut pääoma-arvokertoimet.

Ennusteissa Suomen väestöaineistona käytettiin HMD-tietokannan tasoittamatonta aineistoa vuosilta 1955–2008. Korjauskertoimet määriteltiin yhtiöiltä kerätyn aineiston perusteella. Kuolevuusennustemallina käytettiin Lee-Miller -menetelmää (Lee & Miller (2001)). Laskelmat toteutettiin Rob J. Hyndmanin R-paketilla.

## 2.4 Solvenssi I ja Solvenssi II

Nykyinen Solvenssi I pohjautuu turvaavasti arvioituun vastuuvelkaan, jonka päälle on asetettu pääomavaatimukset yksinkertaisten kaavojen avulla (Sandström (2006)). Solvenssi I ei kuitenkaan enää toteuta nykyajan valvonnan ja riskienhallinnan vaatimuksia. Tätä ongelmaa on Suomen lainsäädännössä korjattu ennakoivan valvonnan kautta (Finanssivalvonnan määräys- ja ohjelmakokoelman luku 11). Solvenssi I:ssä vastuuvelan turvaavuusperiaate on toteutettu käytännössä laskuperusteiden implisiittisten turvaavuusmarginaalien avulla (esimerkiksi käyttämällä erilaisia ikäsiirtoja Makehamin kuolevuusmallissa henki- ja eläkevakuutukseen).

Solvenssi II on rakenteeltaan erilainen. Siinä vastuuvelka lasketaan parhaan estimaatin ja eksplisiittisen riskimarginaalin summana. Pääomavaatimus perustuu kuolevuuden parhaan estimaatin eli odotusarvon ”stressaamiseen” ylös- ja alaspäin. Kalibroinnin tavoitteena on 99.5 % VaR-taso vuoden jaksolla omalle varallisuudelle (Euroopan parlamentin ja neuvoston direktiivi Solvenssi II 2009/138/EY artikla 101).

K2012-raportti pohjautuu parhaan estimaatin mukaiseen kuolevuusennusteeseen ja sen luottamusväleihin. Epävarmuuden arviointi voidaan tehdä analyyttisesti normaalijakauman perusteella (luku 4.4) tai monimutkaisempien simulointimenetelmien avulla (ks. Ronkainen (2012)). Tässä raportissa kuolevuusennnusteesta on jätetty pois kuolevuuden shokkityyppinen kasvu kuten myös Lee ja Carter tutkimuksessaan tekivät (Lee & Carter (1992)). Kyseinen riski koskee erityisesti riskihenkivakuutuksia, ja se tulisi ottaa huomioon vakuutusyhtiön vakavaraisuusvaatimuksessa sekä omassa riski- ja vakavaraisuusarviossa. Kuolevuuden shokkeja on käsitelty muun muassa IAA:n kirjassa ”Stochastic Modeling, Theory and reality from an actuarial perspective” (IAA (2010)).

# 3 Henkivakuutusaineisto

## 3.1 Riskin valinta

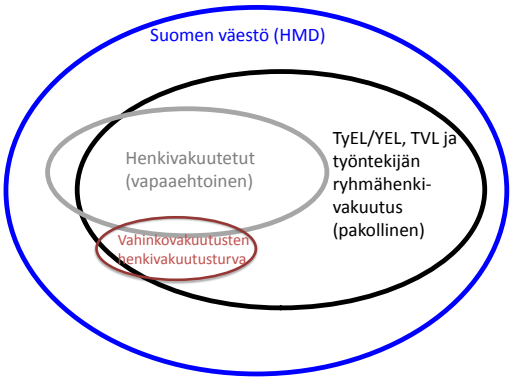
Henkivakuutuksen kannattavuuden kannalta on olennaista, minkälaisia riskejä yhtiö ottaa kantaakseen. Tähän voi vaikuttaa tuotevalikoimalla ja tuoteehdoilla sekä valitsemalla vakuutettavat riskit ja henkilöt, joille vakuutukset myönnetään. Riskit voidaan myös hinnoitella eri henkilöille erilailla.

Henkivakuutetut ovat valikoitunut osajoukko Suomen väestöstä. Henkilöiden tulee olla riittävän terveitä saadakseen vakuutuksen sekä taloudellisesti vakaalla pohjalla selviytyäkseen vakuutusmaksuista. Myös vakuutettujen elintavat poikkeavat keskivertoväestöstä. Havaintojen mukaan vapaaehtoisesti eläke- ja henkivakuutettujen keskimääräinen kuolevuus on matalampi kuin Suomen väestössä keskimäärin. Kuvassa 3.1 on havainnollistettu Suomen väestön, pakollisten vakuutusten (TyEL/YEL ja vahinkovakuutusten henkivakuutusturva sekä lakisääteinen tapaturmavakuutus (TVL) ja siihen liittyvät vakuutukset) ja henkivakuutettujen suhdetta (mittakaavat eivät pidä paikkaansa).

Vapaaehtoisiin henkivakuutuksiin sisältyy paljon erilaisia tuotteita. Vakuutuksen tarve on erilainen eri vakuutetuilla (vrt. lainaturva- ja säästövakuutukset). Myös vakuutuslajien kesken vakuutettujen riskiprofiilit poikkeavat toisistaan. Esimerkiksi säästövakuutettujen riskiprofiili poikkeaa lainavakuutettujen riskiprofiilista. Lisäksi vakuutusyhtiön riski on erilainen eri vakuutuslajeissa. Eläkevakuutuksiin sisältyy negatiivinen kuolinriski, jota nykyisin kutsutaan pitkäikäisyysriskiksi, kun taas henkivakuutuksiin kuolinriski. Säästövakuutuksissa riskisumma määrää vakuutusyhtiön vastuun. Näin ollen muun muassa vastuunvalintakäytännöt sekä turvaavuuskäsitteet eroavat tuotteittain. Tämän takia yksi yhteinen kuolevuus ei välttämättä sovi myöskään kaikille vakuutuslajeille. Tässä raportissa on mahdollisuuksien mukaan laskettu myös vakuutuslajikohtaisia parametreja.

## 3.2 Tutkimusaineisto

Tässä raportissa vakuutusaineistoa on kerätty kaikilta suomalaisilta henkivakuutusyhtiöiltä Suomen liiketoiminnasta vuosilta 2002–2010 (ryhmäeläkevakuutusten osalta pyydettiin tiedot vuosilta 2002–2009). Aineistoa kerättiin kymmenestä eri tuotteesta, jotka näkyvät taulukossa 3.1. Aineisto sisältää vakuutuskohtaista tietoa vakuutuksen alkamisesta ja päättymisestä sekä volyymisuureesta (riskisumma henkivakuutuksessa ja säästöjen määrä eläkevakuutuksissa) sekä mahdollisesta kuolevuuden korotusluokasta (erottelee tapaukset, joilla vakuutusriski on alun perin arvioitu tavallista korkeammaksi).




Kuva 3.1: Henkivakuutettujen suhde Suomen väestössä (ei mittakaavassa)

Taulukko 3.1: Aineiston jakautuminen tuotteittain (%)

| | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|---|---|---|
| Yksilöllinen säästövakuutus | 13.7 | 12.8 | 11.1 | 10.1 | 9.2 | 8.5 | 7.8 | 7.3 | 7.4 |
| Yksilöllinen säästö unit linked | 6.2 | 10.2 | 12.6 | 13.0 | 13.0 | 13.1 | 12.7 | 12.2 | 12.8 |
| Yksilöllinen riskivakuutus | 35.8 | 32.4 | 30.0 | 31.0 | 30.6 | 30.7 | 31.4 | 32.2 | 34.8 |
| Ryhmäetuvakuutus | 8.3 | 8.8 | 7.9 | 7.8 | 9.2 | 9.2 | 9.1 | 9.2 | 10.1 |
| Ryhmäriskivakuutus | 0.9 | 0.8 | 1.2 | 1.1 | 1.1 | 1.1 | 1.0 | 0.9 | 0.8 |
| Yksilöllinen eläkevakuutus | 10.6 | 9.5 | 8.5 | 8.1 | 7.8 | 7.5 | 7.2 | 6.9 | 6.9 |
| Yksilöllinen eläke unit linked | 6.0 | 9.0 | 13.6 | 14.6 | 15.6 | 17.0 | 18.3 | 19.1 | 21.1 |
| Ryhmäeläkevakuutus | 6.4 | 5.9 | 5.5 | 5.6 | 5.6 | 5.5 | 5.5 | 5.5 | 0.2 |
| Perhe-eläke | 0.5 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.0 |
| Tapaturmainen kuolema | 11.5 | 10.2 | 9.1 | 8.3 | 7.5 | 7.0 | 6.6 | 6.2 | 6.0 |
| | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Kuolevuustutkimuksessa tarvitaan tietenkin myös henkilön kuolinpäivämäärä ja syntymäaika. Aineiston tarkempi kuvaus on esitetty liitteessä A.

Alkuperäistä aineistoa on korjattu poistamalla kaikki epäloogiset tietueet tutkimusaineistosta. Havaintoaineistossa oli myös ristiriitaisia kuolinpäivämääriä. Jos kuolinpäivämäärä havaittiin puuttuvan, se lisättiin aineistoon. Jos taas samalla henkilöllä oli useita kuolinpäivämääriä eri vuosina, ne poistettiin havaintoaineistosta. Lisäksi havaintoaineistosta havaittiin, että osalla tietueista niin kutsuttu generoitu henkilötunnus, jonka tarkoituksena oli yksilöidä henkilö, ei ollut yksilöllinen. Nämäkin tapaukset korjattiin aineistoon olemassa olevien tietojen perusteella arvioiden.

Taulukossa 3.1 on esitetty lukumääräjakauma tuotteittain. Aineistossa on eniten yksilöllisiä riskivakuutuksia ja yksilöllisiä eläkevakuutuksia. Vähiten aineistossa on ryhmäriskivakuutuksia, mistä syystä tämä tuote on jouduttu jättämään pois joistakin tulevista tarkasteluista tai se on yhdistetty ryhmäetuvakuutuksiin. Tässä raportissa ryhmävakuutusten kutakin vakuutettua pidetään omana sopimuksenaan.Taulukosta 3.2 on esitetty taulukkoa 3.1 vastaava lukumääräjakauma ilman tuotejaottelua. Taulukossa 3.3 on esitetty vastaava jakauma tuotejaottelulla, jota tullaan käyttämään myöhemmissä tarkasteluissa. On hyvä huomata, ettei taulun 3.1 prosenttiosuuksista päästä summaamalla taulun 3.3 prosenttiosuuksiin. Tämä johtuu siitä, että henkilöllä voi olla useita vakuutuksia eri tuotteissa.

Taulukko 3.2: Aineiston lukumääräjakauma vuosittain

| | Lkm |
|---|---|
| 2002 | 1 546 892 |
| 2003 | 1 713 616 |
| 2004 | 1 877 782 |
| 2005 | 1 955 007 |
| 2006 | 2 028 341 |
| 2007 | 2 065 769 |
| 2008 | 2 115 050 |
| 2009 | 2 134 959 |
| 2010 | 1 989 039 |

Taulukko 3.3: Jakauma tuotteittain ryhmitellyssä aineistossa (%)

| | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | Yht |
|---|---|---|---|---|---|---|---|---|---|---|
| Säästövakuutus | 20 | 23 | 23 | 23 | 22 | 21 | 20 | 19 | 20 | 21 |
| Yksilöllinen riskivakuutus | 36 | 33 | 31 | 32 | 31 | 31 | 32 | 33 | 35 | 33 |
| Ryhmäetu- ja ryhmäriski | 9 | 10 | 9 | 9 | 10 | 10 | 10 | 10 | 11 | 10 |
| Yksilöllinen eläke ja perhe-eläke | 17 | 19 | 22 | 23 | 23 | 24 | 25 | 26 | 27 | 23 |
| Ryhmäeläkevakuutus | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 0 | 5 |
| Tapaturmainen kuolema | 12 | 10 | 9 | 8 | 8 | 7 | 7 | 6 | 6 | 8 |
| | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | |

## 3.3 Kohortin koon vaikutus

Kuolevuusperusteet perustuvat yleensä oletukselle, että vakuutuskohortit ovat niin suuria, että äärellisestä otoskoosta aiheutuva hajautuva riski on riittävän pieni, jotta se voidaan jättää huomiotta (suurten lukujen laki). Pienissä vakuutuskannoissa tai sen osissa tämä oletus ei kuitenkaan päde. Tässä raportissa nähdään satunnaisvaihtelun merkitys myöhemmissä luvuissa (luku 5). Kohortin koon vaikutus voidaan ottaa eksplisiittisesti huomioon joko hajonnan avulla (Gerber (1990)) tai simuloimalla (Ronkainen (2012)). Viimeksi mainitussa lähteessä on esimerkkejä eläkkeen stokastisten pääoma-arvojen laskennasta kohortin koon vaihdellessa. Alho (2005) on tarkastellut asiaa teoreettisesti binomi- ja Poisson-jakaumien avulla.

Ei-hajautuva eli systemaattinen riski koskee kaikkein suurimpiakin vakuutuskantoja, ja se tulee aina ottaa huomioon kuolevuusmallinnuksessa. Tässä raportissa on käytetty stokastista Lee-Miller -mallia parhaan estimaatin kuolevuusennusteessa sekä luottamusvälien arvioinnissa eikä otoskoon vaikutusta ole otettu huomioon.

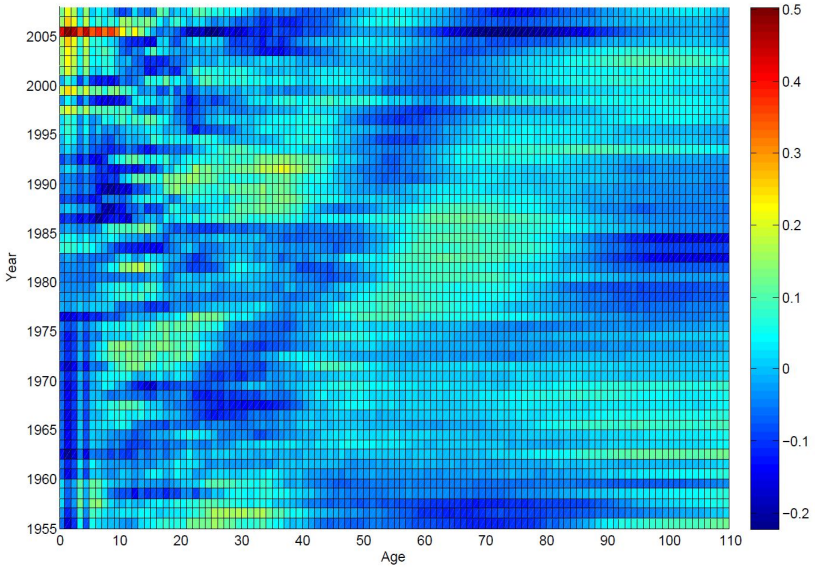
## 3.4 Vastuunvalinnan vaikutus

Eräillä henkilöryhmillä saattaa olla muita henkilöryhmiä suurempi intressi ottaa henkivakuutus, koska he kokevat riskinsä kuolla suurentuneen. Henkilölle, jonka elinajanodote lääkärinlausunnon mukaan on enää viikkoja, ei vakuutusta ymmärrettävästi myönnetä. Kuitenkin jo pienempikin kuolinriskin kasvaminen voi vaikuttaa mahdollisuuteen saada vakuutus tai vakuutus myönnetään vain korotetuin maksuin.

Yhtiö arvioi vastuunvalinnan tarpeellisuuden ja laajuuden tapauskohtaisesti. Tähän vaikuttavat vakuutustuote ja riskin suuruus. Terveysselvitys voi

sisältää vain työkykyisyysilmoituksen tai se voi sisältää laajoja lääketieteellisiä tutkimuksia. Pitkäikäisyysriskiä ei arvioida vastuunvalinnalla.

Hyvä vastuunvalinta vaikuttaa osaltaan siihen, että vakuutettujen kuolevuus poikkeaa väestön kuolevuudesta. Seppo Koskinen on todennut, että valinnan vaikutus (selekti) keskittyy vakuutusajan alkuvuosiin ja kuolleisuuteen sillä on vaikutusta noin kymmenen ensimmäisen vuoden ajan (SHY (1982) s. 25). Myös vakuutusyhtiöiden riskiperusteanalyyseissä havaittu kuolevuus kahdelta ensimmäiseltä vakuutusvuodelta raportoitiin erikseen. Toisaalta Vesa Ronkainen on havainnut, että sotavuosina syntyneiden kohortin kuolleisuus on pysynyt koko ajan pienempänä kuin muiden (kuva 3.2, Ronkainen (2012)).




Kuva 3.2: Ronkainen (2012) kuva 4.15, jossa sotavuosina syntyneen kohortin pienempi kuolevuus näkyy tummansinisenä.

Toisin sanoen vastuunvalinnalla on suuri merkitys myös kuolevuuteen. Vastuunvalinta selittää osittain henkivakuutusaineiston ja populaatioaineiston kuolevuustasojen eroja. Vastuunvalinnan vaikutusta ei kuitenkaan ole tässä raportissa voitu analysoida.

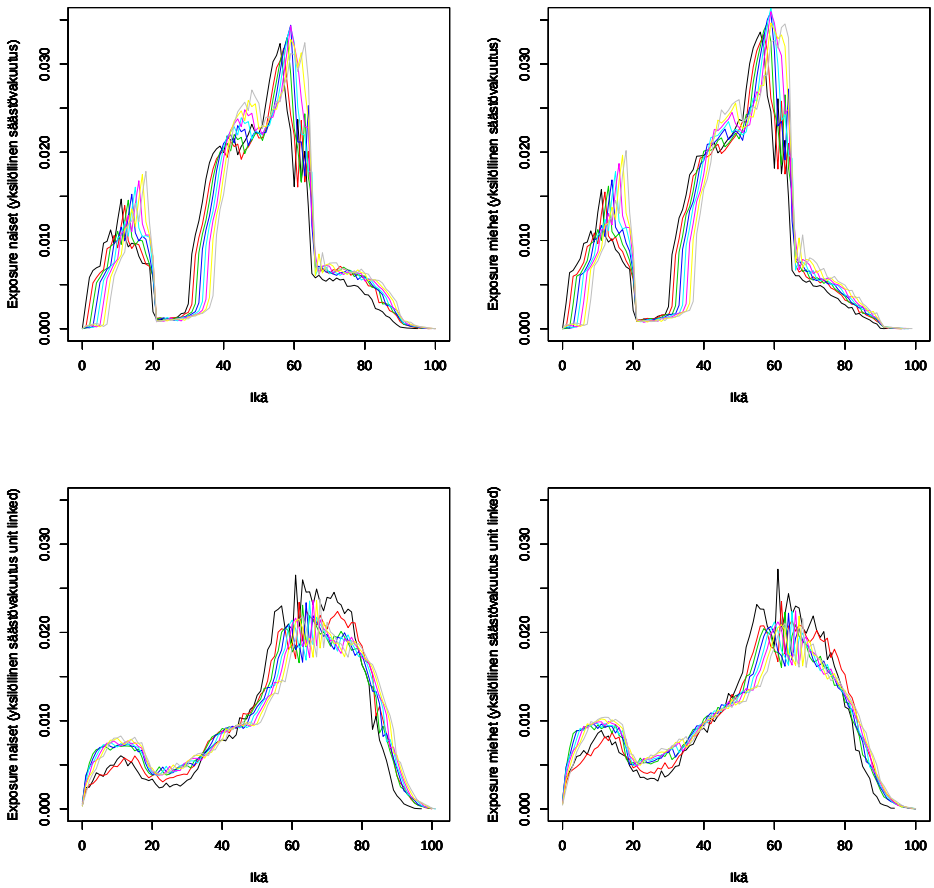
## 3.5 Säästöhenkivakuutukset

Säästöhenkivakuutukset ovat vakuutuksia elämisen varalta. Toisin sanoen tarkoituksena on kerryttää vakuutussäästöjä vakuutussopimuksessa määritellyn edunsaajan myöhempään käyttöön. Puhtaassa säästöhenkivakuutuksessa osan vakuutetuista oletetaan kuolevan ja näiden vastuiden vapautuvan muille vakuutuksille niin kutsuttuna kuolevuushyvityksenä.

Kuitenkin säännöllisesti vakuutukseen sisällytetään turva myös kuoleman varalta. Jos säästösumma ja kuolemanvaraturva ovat yhtä suuret, vakuutusyhtiöllä ei ole vakuutusriskiä. Aikaisemmin tällainen ei ollut sallittua, vaan vakuutuksen tuli sisältää vakuutusmaksusta laskettavaa vakuutusriskiä vähintään 3 % tai vakuutussummasta laskettava 5 %:n riski.

Tässä raportissa yksilöllisiin säästövakuutuksiin kuuluvat perustekorkoiset ja sijoitussidonnaiset (unit link) vakuutukset sekä näiden yhdistelmät eli hybrideituotteet. Viime vuosina sijoitussidonnaisten tuotteiden osuus on kasvanut huomattavasti (taulukko 3.1).

Kuvassa 3.3 on esitetty kuolinriskille altistuneiden (exposure) ikäjakauma vuosittain (altistuma iässä $x$ vuonna $t$ / altistuma vuonna $t$). Kuvassa kukin väri edustaa yhtä vuosista 2002–2010. Huomataan, että suurin osa lapsille otetuista säästöhenkivakuutuksista lunastetaan täysi-ikäiseksi tultaessa. Uudelleen säästöhenkivakuutusten suosio kasvaa noin 40 vuoden iässä. Lisäksi huomataan, että suurin osa perustekorkoisista tuotteista raukeaa eläkeiän tienoilla, kun taas sijoitussidonnaisia vakuutuksia pidetään pidempään.



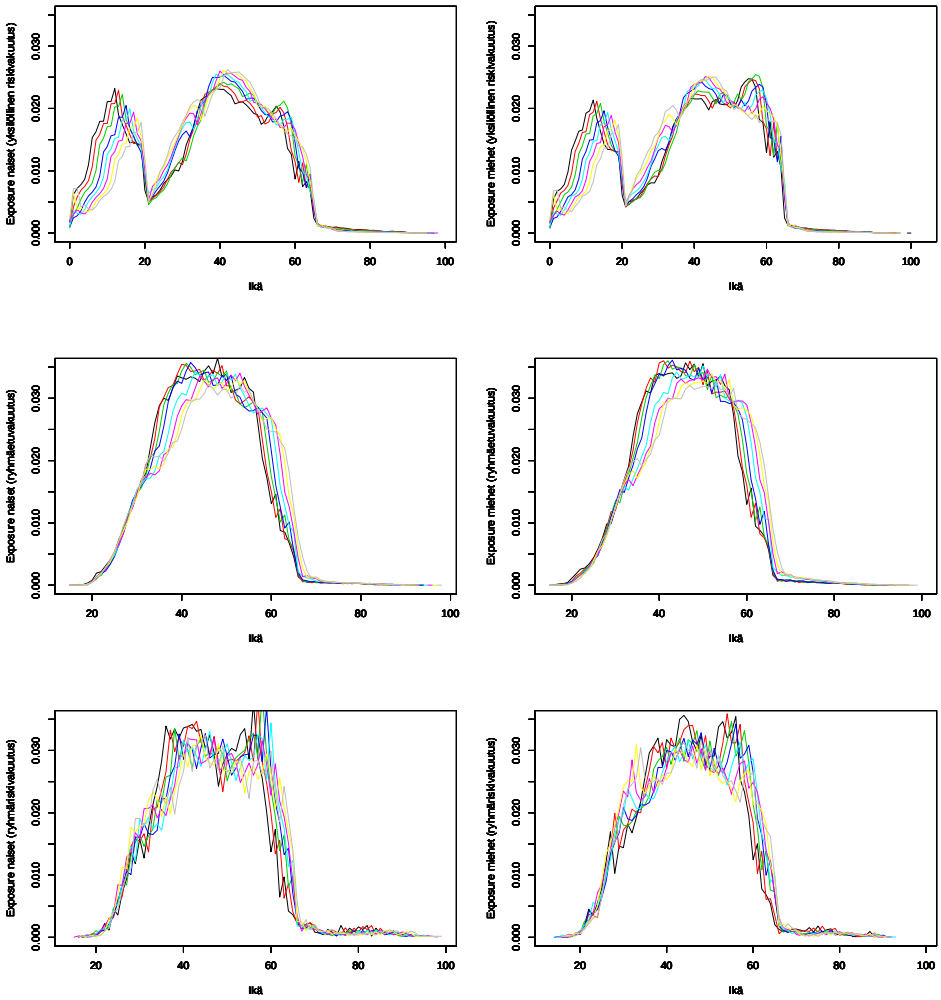

Kuva 3.3: Säästöhenkivakuutusten ikäjakaumat vuosittain (eri väriset viivat kuvaavat eri vuosia)

## 3.6 Riskivakuutukset

Riskivakuutus otetaan kuoleman varalta yleensä tietylle ajanjaksolle. Esimerkiksi lainaturvat räätälöidään lainapääoman ja laina-ajan perusteella. Vastuunvalinnassa otetaan huomioon muun muassa ikä ja terveydentila. Jos so-

pimus sidotaan pitkäksi aikaa, vastuunvalinnan vaikutus kuolevuuteen vähenee vakuutuksen vanhetessa.

Suurin osa yksilöllisistä riskivakuutuksista päättyy eläkeikään. Ryhmäetuvakuutuksissa kohderyhmä on nuorista aikuisista reiluun 60 ikävuoteen. Ryhmäriskivakuutuksia aineistossa on vähän, mikä näkyy myös kuvaajassa 3.4 vaihteluna.

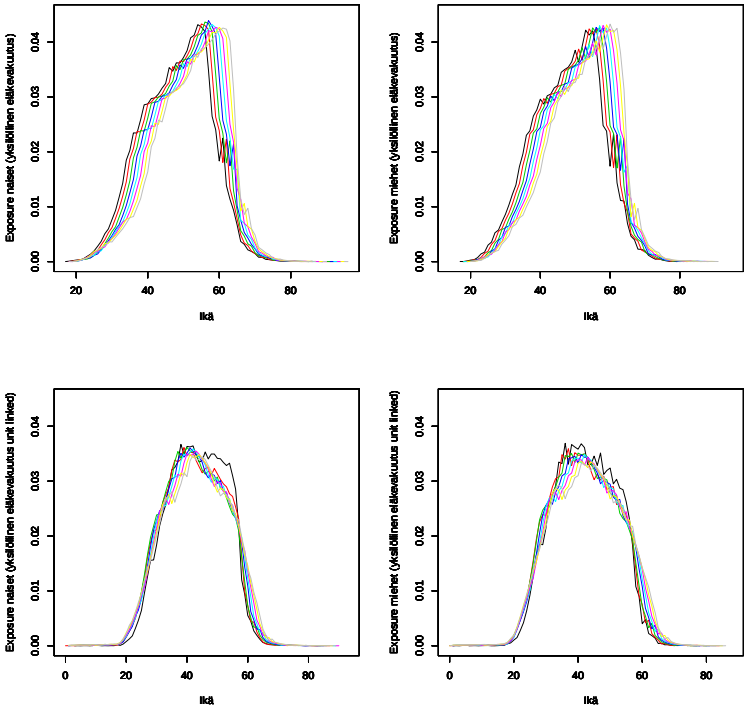



Kuva 3.4: Riskivakuutusten ikäjakaumat

## 3.7 Eläkevakuutukset

Eläkevakuutuksessa vakuutusyhtiölle aiheutuu riskiä siitä, että vakuutettu kohortti elää pidempään kuin on oletettu (systemaattinen eli ei-hajautuva riski). Myös kohortin koolla on vaikutusta riskiin, kuten edellä kuvattiin (hajautuva riski, luku 3.3). Keskeinen tekijä on myös eläkevakuutuksen kesto. Useimmat yksilölliset eläkevakuutukset ovat määräaikaisia, eläkeaika voi olla esim. ikäväli 63–68 (alin eläkeikä vaihtelee eri aikoina lainsäädännön muutosten vuoksi). Tällöin pitkäikäisyysriskin merkitys on rajoitettu verrattuna

elinikäisen eläkkeen tapaukseen, jollaisia ovat useat vanhemmat ryhmäeläkkeet eli ns. vapaamuotoiset eläkkeet (vpm). Nämä erot käyvät ilmi kuvista 3.5–3.6. Luonnollisesti myös riskisumma vaikuttaa yhtiön taloudelliseen riskiin, kuten aiemmin on todettu.




Kuva 3.5: Eläkevakuutusten ikäjakaumat

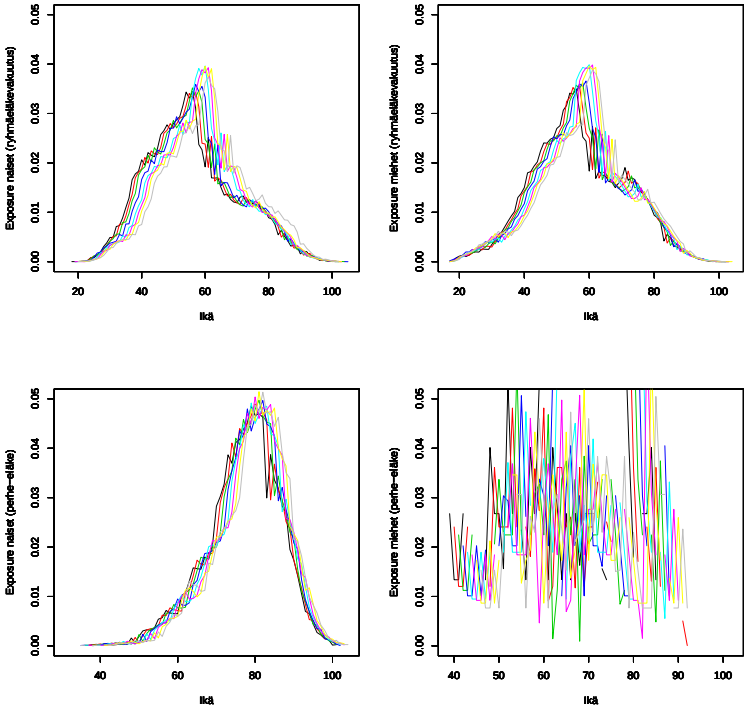
## 3.8 Perhe-eläke

Perhe-eläkkeeseen sisältyy myös turva vakuutetun kuoleman varalta. Sen sijaan, että korvaus maksettaisiin kertakorvauksena, perhe-eläkkeessä korvaus maksetaan eläkkeenä sovittuna ajanjaksona tai edunsaajan eliniän.

Kuten taulukossa 3.1 näkyy, yksilöllisen perhe-eläkkeen aineiston suppeuden vuoksi sitä ei voi itsenäisesti analysoida. Erityisesti miehillä havaintoja on todella vähän, mikä näkyy kuvassa 3.6 vaihteluna. Näin ollen perhe-eläke yhdistetään yksilöllisten eläkevakuutusten kanssa.

## 3.9 Tapaturmainen kuolema

Henkivakuutusyhtiöt eivät yleensä myy tapaturmavakuutuksia itsenäisinä tuotteina, vaan ne liittyvät johonkin muuhun myytyyn henkivakuutustuotteeseen. Tuotteen myynti on vähentynyt, mikä näkyy taulukoista 3.1 ja 3.3.




Kuva 3.6: Perhe- ja ryhmäeläkevakuutusten ikäjakaumat summattuna

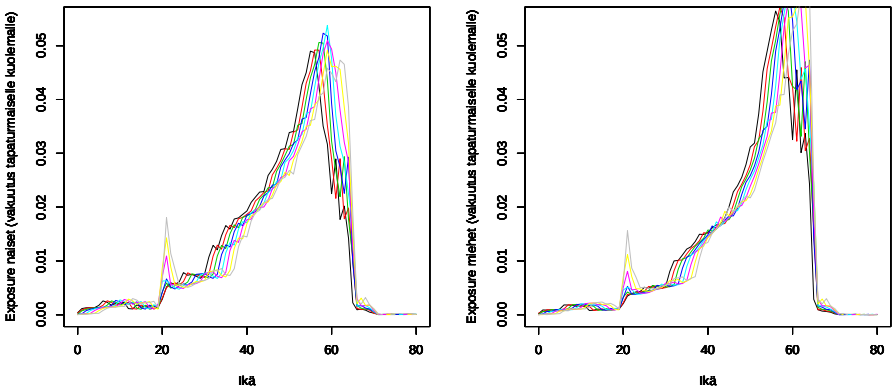
## 3.10 Yhteenveto

Vakuutustuotteet ovat painottuneet eri ikäväleille seuraavasti:

- Säästövakuutuksissa havainnot ovat keskittyneet alle 20-vuotiaille ja yli 35-vuotiaasta noin 65-vuotiaaseen. Ikävälillä 20–35 havaintoja on vähemmän. Säästövakuutuksia löytyy iäkkäämmiltäkin.
- Riskivakuutuksia löytyy erityisesti ikäväliltä 30–60 vuotta (lainavakuutukset). Havaintoja on myös iäkkäämmiltä.
- Eläkevakuutuksissa havaintoja on erityisesti ikäväliltä 35–65 vuotta. Ryhmäeläkevakuutuksissa havaintoja on myös iäkkäämmiltä.
- Tapaturmaisessa kuolemanvaravakuutuksessa havaintoja on ikäväliltä 30–65 vuotta.

Henkivakuutusaineistoa on niukasti nuorimmilla ja vanhimmilla ikäluokilla. Tämä rajoittaa aineiston analyysiä kuten jäljempänä käy ilmi.

Vakuutustuotteiden kuolevuusriskit poikkeavat toisistaan. Vakuutussopimukset poikkeavat toisistaan myös vakuutustuotteen sisällä muun muassa voimassaoloajan ja raukeamisen osalta. Lisäksi vakuutuskannan rakenteessa on eroja vakuutusyhtiöiden välillä. Näin ollen myös henkivakuutusaineisto on heterogeeninen.




Kuva 3.7: Tapaturmavakuutusten ikäjakauma

# 4 Väestökuolevuuden ennustaminen

## 4.1 Yleistä

Vakuutusyhtiöaineiston suppeuden vuoksi henkivakuutettujen kuolevuuden pitkän ajan ennuste ei voi pohjautua ainoastaan henkivakuutusyhtiöiltä kerättyyn aineistoon. Näin ollen uusi henkivakuutusyhtiöiden referenssikuolevuus pohjautuu Suomen väestön kuolevuusaineistoon (saatavilla Human Mortality -tietokannasta osoitteesta `www.mortality.org`), jonka päälle on määritelty korjauskerroin väestökuolevuuden ja henkivakuutuskuolevuuden välille. Näin saadaan määriteltyä henkivakuutettujen kuolevuus. Tässä luvussa keskitytään kuolevuuden ennustamiseen Suomen väestöaineistolla (HMD).

## 4.2 Menetelmän valinta

Edellisessä henkivakuutusyhtiöiden referenssikuolevuustutkimuksessa K2004-väestöennuste tehtiin Lee-Carter -mallilla. Lee-Carter -malli (Lee & Carter (1992)) ja sen muunnelmat ovat edelleen laajasti käytettyjä kuolevuuden ennustemalleja. Lee-Carter -mallissa keskikuolevuus $m_{x,t}$ iässä $x$ vuonna $t$ ($t \in [s, T]$ =havaintoperiodi) mallinnetaan seuraavasti

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}, \tag{4.1}$$

missä ln on luonnollinen logaritmi, $m_{x,t}$ on kuolleiden lukumäärä suhteessa altistumaan (ks. HMD-tietokannan protokolla `www.mortality.org`) ja $\epsilon_{x,t}$ normaalijakautunut virhetermi. Parametri $a_x$ kuvaa kuolevuuden keskimääräistä tasoa ja $b_x$ kuolevuuden muutoksen intensiteettiä iässä $x$. Parametri $k_t$ kuvaa kuolevuuden suhteellista tasoa vuonna $t$.

Lee-Carter-malli on yli-identifioituva, joten sitä on rajoitettu seuraavilla ehdoilla:

$$\sum_t k_t = 0 \text{ ja } \sum_x b_x = 1,$$

missä $t$ viittaa lähtöaineiston vuosiin ja $x$ ikiin. Rajoitteesta seuraa (Alho et al. (2012)), että estimaatti parametrille $a_x$ saadaan seuraavasti

$$\hat{a}_x = \frac{1}{T - s + 1} \sum_{t=s}^{T} \ln(m_{x,t}). \tag{4.2}$$

Parametrit $b_x$ ja $k_t$ voidaan estimoida muun muassa singulaariarvohajotelmalla (SVD). Lee-Carterin alkuperäisessä artikkelissa parametrin $k_t$ valinnassa käytettiin lisäksi ehtoa, että havaintoaineistosta havaittujen kuolleiden

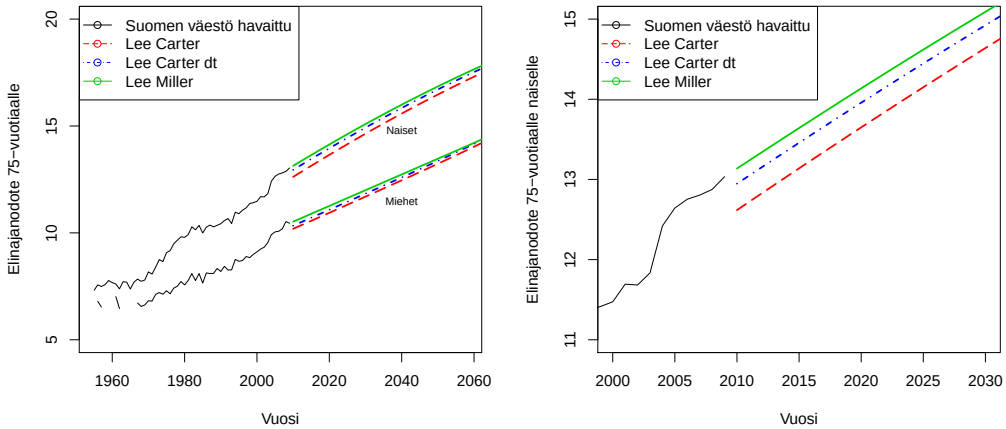
lukumäärä $D_t$ vastaa mallilla estimoitua kuolleiden lukumäärää vastaavina vuosina. Tätä ehtoa ei ole kuitenkaan otettu huomioon K2004-tutkimuksessa. Näin ollen tässä raportissa Lee-Carter -menetelmällä viitataan menetelmään, jossa parametrit on estimoitu singulaariarvohajotelmalla ottamatta huomioon kuolleiden lukumäärään liittyvää ehtoa, ellei toisin mainita.

Ennuste perustuu parametrin $k_t$ autoregressiiviseen aikasarjaan. Yleisesti on havaittu, että aikasarjamalli "random walk with drift" sopii useimmissa tapauksissa lähtöaineistoon (Lee & Carter (1992)). Tässä raportissa käytetään samaa muotoa. Toisin sanoen oletamme, että

$$k_t = k_{t-1} + c + e_t, \tag{4.3}$$

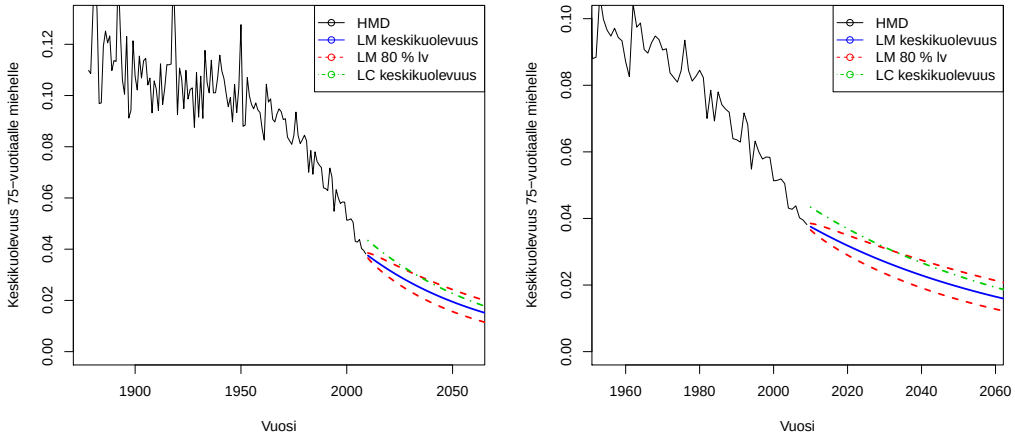
missä $t = T+1, ..., T+h$, $c$ on vakio ("drift"), $e_t$ normaalijakautunut virhetermi ja $h$ =ennustevuosien lukumäärä. Vakiotermin estimaatti on muotoa $\hat{c} = \frac{k_T - k_s}{T-s}$, missä $T - s$ = lähtöaineiston pituus vuosina $-1$.

Lee-Carter -menetelmästä on esitetty paljon erilaisia muunnelmia. Muun muassa Booth et al. (2006) vertailivat artikkelissaan yleisemmin käytettyjä Lee-Carterin muunnelmia ja totesivat, että vaikka muunnelmat ovat usein parempia logaritmoitujen keskikuolevuuksien ennustamisessa, ne eivät välttämättä ole parempia elinajanodotteiden ennustamisessa. Vertailu oli tehty Lee-Carter -menetelmään, jossa on otettu huomioon kuolleiden lukumäärään liittyvä ehto (tähän menetelmään viitataan merkinnällä "Lee-Carter dt"). Kuvassa 4.1 on havainnollistettu Lee-Carter -menetelmästä aiheutuvaa virhettä 75-vuotiaan elinajanodotteissa. Huomataan, että virhe on suurinta Lee-Carter -ennusteessa, jossa ei ole otettu huomioon kuolleiden lukumäärään liittyvää ehtoa. Elinajanodotteet, joissa on jatkettu empiiristä aikasarjaa, on laskettu yhdenmukaisuussyistä periodeittain. Kuvassa 4.2 on 75-vuotiaan miehen keskikuolevuudet Lee-Carter ja Lee-Miller -menetelmillä. Virhe voi olla positiivinen tai negatiivinen havaintoaineistosta riippuen.




Kuva 4.1: 75-vuotiaiden naisten ja miesten elinajanodotteet Lee-Carter ja Lee-Miller -menetelmillä

Leen ja Carterin artikkelissa esittämä korjaus liittyen kuolleiden lukumäärään $D_t$ ei ole riittävä korjaamaan koko virhettä, joka Lee-Carter -mallissa




Kuva 4.2: 75-vuotiaan miehen keskikuolevuudet Lee-Carter ja Lee-Miller -menetelmillä

syntyy. On nimittäin osoitettu, että Lee-Carter -mallissa syntyy ns. ”jump-off”-virhe. Syynä virheelle on se, että Lee-Carter -mallissa ennuste lähtee sovitetuista arvoista eikä todellisista havaituista kuolevuuksista. Tästä syystä viimeisen havaitun arvon ja ensimmäisen ennustetun arvon välille syntyy hyppy, joka voi olla positiivinen tai negatiivinen. Koska Lee-Carterin alkuperäiseen menetelmään liittyy edellä mainittu virhe, tässä raportissa tullaan käyttämään kuolevuusennusteessa eri menetelmää kuin K2004-mallissa.

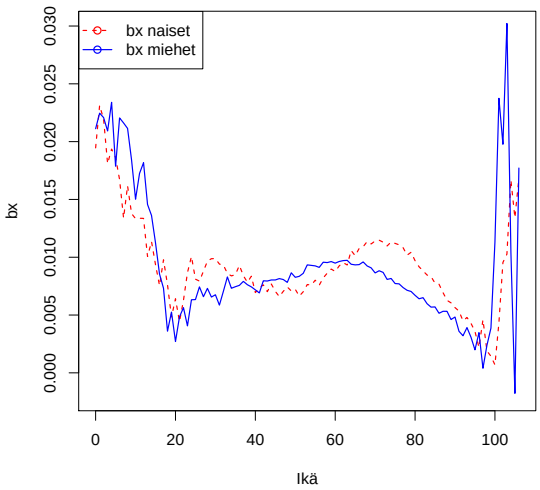
Muunnelmia on lukuisia ja useimmat niistä korjaavat Lee-Carterissa havaitun ”jump-off”-virheen. Muun muassa Lee-Miller -muunnelmassa virhe otetaan huomioon siten, että ennuste lähtee viimeisestä havaitusta arvosta eikä ennusteen parametrien mukaan lasketusta arvosta. Lisäksi parametri $k_t$ valitaan siten, että havaitut vastasyntyneen periodikohtaiset elinajanodotteet täsmäävät mallin sovitteesta laskettuihin elinajanodotteisiin (Lee & Miller (2001)). Booth-Maindonald-Smith valitsee $k_t$:n hyvyyskriteerin perusteella (”goodnees of fit”), kun $k_t$ oletetaan lineaariseksi. Periaate on hyvä teoriassa, mutta käytännössä menetelmä näyttää antavan liian lyhyitä aikaperiodeja pitkän aikavälin ennusteissa. Hyndman ja Ullah kehittivät yleisemmän menetelmän, jossa muun muassa yleisimmät aikasarjamallit ovat käytettävissä. Lisäksi kuolevuudet estimoidaan ei-parametrisilla tasoitusmenetelmillä. Booth et al. (2006) artikkelissa on tehty hyvä vertailu muun muassa näiden menetelmien välillä.

Tässä raportissa valittiin menetelmäksi Lee-Miller, sillä se on yksinkertainen menetelmä, joka ottaa huomioon Lee-Carter -mallin suurimmat puutteet eikä monimutkaisemmilla muunnelmilla saavuteta merkittävästi parempia tuloksia (Booth et al. (2006)). Lee-Miller -menetelmää käytettiin myös lakisääteisen tapaturmavakuutuksen ja liikennevakuutuksen kuolevuusreferenssimallissa (PricewaterhouseCoopers (2011)). Tästä eteenpäin raportissa esitetyt tulokset on tehty Lee-Miller -mallilla.

## 4.3 Havaintoaineisto

### 4.3.1 Lähtöaineiston valinta

HMD-tietokannasta on saatavilla Suomen väestön kuolevuudet tasoittamattomina ja tasoitettuina (periodikuolevuus). Tasoittamattomassa Suomen väestöaineistossa on havaittavissa satunnaisuutta sekä puuttuvia ja poikkeavia havaintoja erityisesti vanhoissa ikäluokissa. Tämä johtuu populaation ja kuolleiden lukumäärän pienuudesta. Aineiston puutteet vaikuttavat parametrien estimointiin, mikä näkyy myös kuolevuusennusteessa (kuva 4.3). Referenssikuolevuuden ei tulisi sisältää vaikeasti tulkittavia tai epäloogisia arvoja.




Kuva 4.3: Tasoittamattoman aineiston parametri $b_x$ naisille ja miehille

Lähtöaineiston ikäväliä ei referenssikuolevuudessa voida rajoittaa. Tämän vuoksi tutkimuksessa käytetään HMD-tietokannan tasoitettua lähtöaineistoa. Tasoitus on tehty vanhemmille ikäluokille. Nuorten ikäluokkien nollakuolevuudet on Hyndmanin R-paketissa interpoloitu saman ikäryhmän viereisten vuosien arvoilla. Lee-Miller -mallin parametreja ei ole tämän jälkeen enää erikseen tasoitettu.

Toinen tapa on käyttää lähtöaineistona rajoitettua tasoittamatonta aineistoa ja tehdä tasoitus kuolevuusennusteelle asiantuntijaharkinnalla (PricewaterhouseCoopers (2011)) tai jollakin tasoitusmenetelmällä.

### 4.3.2 Lähtövuosien valinta

Lähtöaineiston valinnalla on hyvin merkittävä vaikutus ennusteeseen. Lee-Miller -kuolevuusennusteen trendin kulmakerroin määräytyy aikasarjamallin vakiokertoimen perusteella ($\hat{c} = \frac{k_T - k_s}{T - s}$). Koska kulmakerroin määräytyy havaintoaineiston viimeisen ja ensimmäisen vuoden perusteella, lähtövuosien valinta on ennusteen kannalta yksi tärkeimmistä päätöksistä. Käytännössä sen valinnassa tarvitaan asiantuntijaharkintaa.

Lähtövuosien valinnassa tulee ottaa huomioon, että tässä raportissa tavoitteena on tehdä pitkän aikavälin ennuste. Lee-Miller -menetelmä on tilastollinen menetelmä, jonka luotettava estimointi vaatii riittävän pitkän aineiston. Lisäksi hajonta on tyypillisesti suurempaa pitkässä aineistossa, mikä heijastuu ennusteisiin leveämpinä luottamusväleinä. Lyhyessä aineistossa ennusteen epävarmuus saattaa tulla aliarvioiduksi. Lähtöaineiston tulisi siis olla riittävän pitkä.

Toisaalta todella pitkässä lähtöaineistossa ongelmana on se, että kuolevuuden kehitys eri ikäluokissa on muuttunut yli ajan eikä välttämättä ole realistista olettaa samanlaisen kehityksen jatkuvan tulevaisuudessa. Ennustetta tehtäessä halutaankin ottaa huomioon se viimeisin kuolevuuden kehitys, jonka voidaan uskoa jatkuvan tulevaisuudessa. Näin ollen mahdollisimman pitkä lähtöaineisto ei välttämättä ole sopivin.

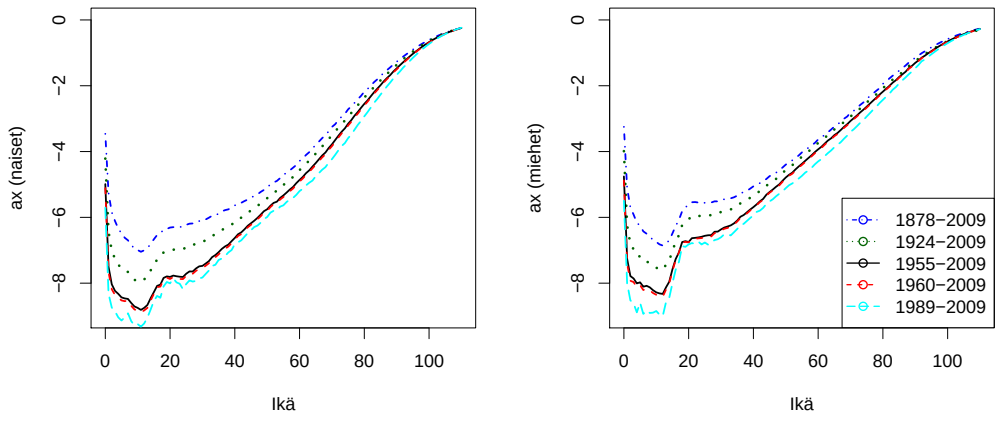
Seuraavissa tarkasteluissa tutkitaan Suomen väestöaineistolla (HMD), mikä aikaperiodi olisi sopivin henkivakuutusyhtiöiden referenssikuolevuuden lähtöaineistoksi. Tarkasteluissa käytetään samoja dataperiodeja kuin Sillan raportissa (Kuusela & Kukkala (2010)) lisättynä K2004-tutkimusta vastaavalla aikaperiodilla 1960–2009. Dataperiodit ovat siis seuraavat:

- 1878–2009
- 1924–2009
- 1955–2009 (lakisääteisen tapaturmavakuutuksen ja liikennevakuutuksen kuolevuusreferenssimallissa käytetty aikaperiodi)
- 1960–2009 (vrt. K2004 1960–2001)
- 1989–2009

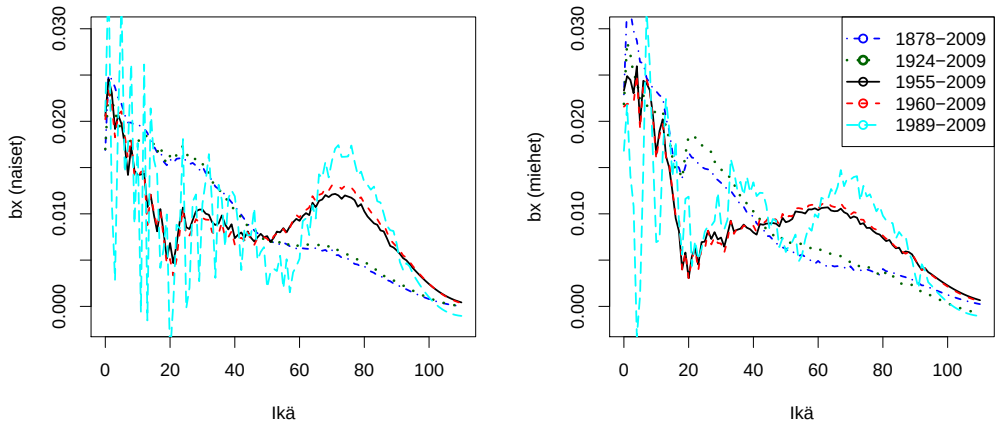
Ensimmäiset kaksi aikaperiodia edustavat mahdollisimman pitkiä aikavälejä, kun taas loput on valittu siten, että sotien vaikutusta ei olisi enää aineistossa. Lyhin dataperiodi 1989–2009 vastaa Booth-Maindonald-Smith -mallin valitsemaa aikaväliä, kun pienin mahdollinen aikaväli rajoitetaan 20 vuoteen (Hakkarainen (2012)). Aineistoja ei ole tasoitettu seuraavissa tarkasteluissa (Sillan raportissa tehtiin tasoitus miehille aikaperiodilla 1924–2009).

Aloitetaan tarkastelu Lee-Miller -mallin parametreja tutkimalla. Kuvassa 4.4 on esitetty kuolevuuden keskimääräinen taso $a_x$ iässä $x$ (ks. kaava 4.2). On luonnollista, että mitä pidempi aikaväli on valittu, sitä suurempia parametrin $a_x$ arvot ovat. Parametri $a_x$ ei riipu vuodesta $t$, joten sen oletetaan pysyvän samana koko ennustejakson ajan.

Kuolevuuden vähenemisen intensiteettiä iässä $x$ kuvaava parametri $b_x$ on esitetty kuvassa 4.5. Huomataan, että pitkillä aikaperiodeilla (1878–2009 ja 1924–2009) parametri $b_x$ on suurimmillaan nuorissa ikäluokissa. Näin ollen näillä aikaperiodeilla parametri $k_t$ eli kuolevuuden suhteellinen taso (kaava 4.1) tulee painotetuksi eniten nuorimmissa ikäluokissa. Lyhyemmillä aikaperiodeilla painotus on siirtynyt vanhempiin ikäluokkiin. Tämä on odotettuakin, sillä kuolevuuden aleneva kehitys on siirtynyt vanhempiin ikäluokkiin. Huomataan myös, että lyhimmällä aikaperiodilla 1989–2009 vaihtelu parametrissa $b_x$ on todella suurta. Tämä viittaa siihen, että aikaperiodi on liian lyhyt.




Kuva 4.4: Parametri $a_x$ naisille ja miehille

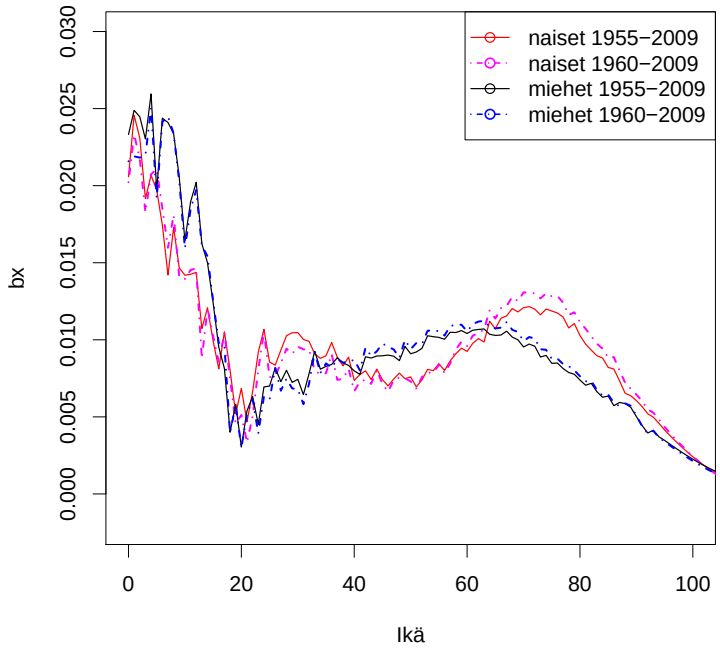



Kuva 4.5: Parametri $b_x$ naisille ja miehille

Kuvassa 4.6 on asetettu aikaperiodit 1955–2009 ja 1960–2009 samaan kuvaan erikseen miehille ja naisille. Huomataan, ettei parametri $b_x$ eroa paljon näillä aikaperiodeilla. Mielenkiintoista on kuitenkin huomata, että aikaperiodilla 1960–2009 naisilla eroa on nimenomaan nuorilla aikuisilla ja iäkkäämmillä. Kehitys vaikuttaa loogiselta. Ero vaikuttaa kuitenkin melko pieneltä, joten todennäköisesti sillä ei ole vaikutusta kuolevuuteen tai elinajanodotteisiin. Lisäksi huomataan, että naisilla ja miehillä parametrin $b_x$ arvot eroavat eri ikäluokissa.

Edellä esitettyjen havaintojen perusteella voidaan arvioida, ettei parametrin $b_x$ taso ole vakio pitkällä aikavälillä. Kuitenkin Lee-Miller -mallissa oletetaan, että parametri $b_x$ pysyy vakiona koko ennusteperiodin ajan. Tätä ominaisuutta voitaisiin parantaa muun muassa lokaaleilla malleilla, mitä on testattu esimerkiksi Suomen aineistolla (Ronkainen (2012) luku 4.6). Mainitussa tutkimuksessa havaittiin, että otettaessa huomioon kohdassa 3.4 mainittu sota-aikana syntyneen kohortin alempi kuolevuus ei monimutkaisemmalla

mallilla kuitenkaan saatu merkittävästi parempaa kuolevuuden ennustetta.




Kuva 4.6: Parametri $b_x$ naisille ja miehille aikaperiodeilla 1955–2009 ja 1960–2009

Kuvassa 4.7 on esitetty kuolevuuden suhteellinen taso vuonna $t$ eli parametri $k_t$. Huomataan, että pitkillä aikaperiodeilla parametri $k_t$ ei ole lineaarinen. Ennusteen kannalta on mielenkiintoista tietää, miten ennusteen vakiotermi eli kulmakerroin muuttuu eri aikaperiodeilla (kaava 4.3). Aikaperiodilla 1924–2009 vakiotermi vaikuttaa olevan pienimmillään. Vertailtaessa aikaperiodeja 1955–2009 ja 1960–2009 estimaattorit ($\hat{c} = \frac{k_T - k_s}{T-s}$) ovat seuraavat:

1955–2009 naiset: $\hat{c} = -2.17$

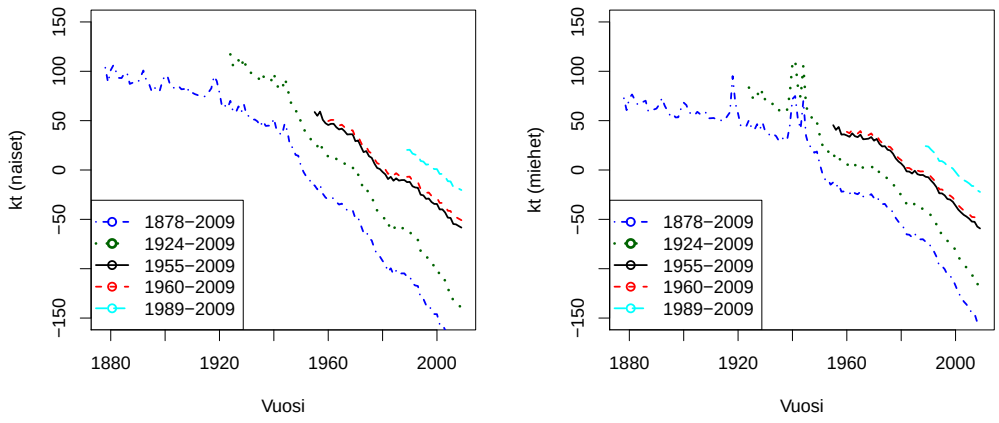
1960–2009 naiset: $\hat{c} = -2.05$

1955–2009 miehet: $\hat{c} = -1.94$

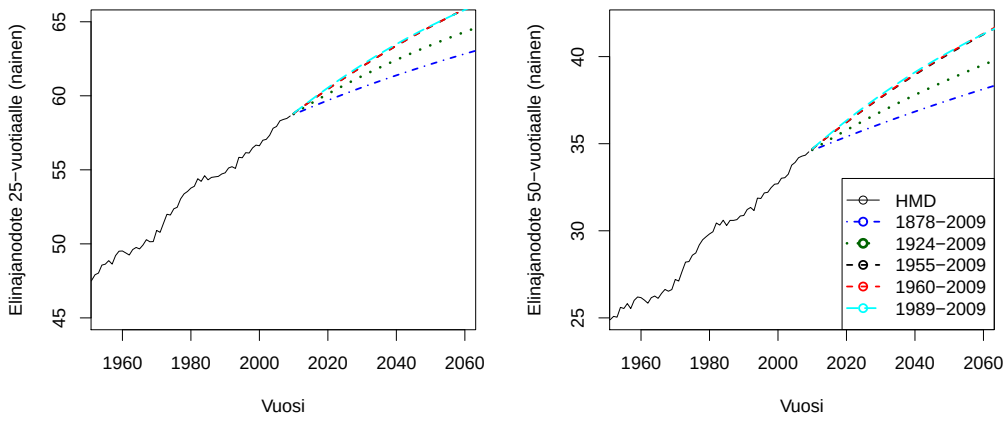
1960–2009 miehet: $\hat{c} = -1.90$

Näin ollen suurta eroa vakiossa $\hat{c}$ ei ole, mutta lyhyempi aikaperiodi antaa hiukan loivemman ennusteen. Tämä voi kuitenkin johtua vain satunnaisvaihtelusta.

Tutkitaan seuraavaksi, minkälainen vaikutus eri aikaperiodeilla on elinajanodotteisiin ja keskikuolevuuksiin (kuviot 4.8–4.11). Liitteen B taulukossa on esitetty sukupuolittain elinajanodotteet 0-, 30-, 50- ja 70-vuotiaalle vuosina 2010, 2015, 2020, 2030 ja 2050 (elinajanodotteet laskettu periodeittain).




Kuva 4.7: Parametri $k_t$ naisille ja miehille

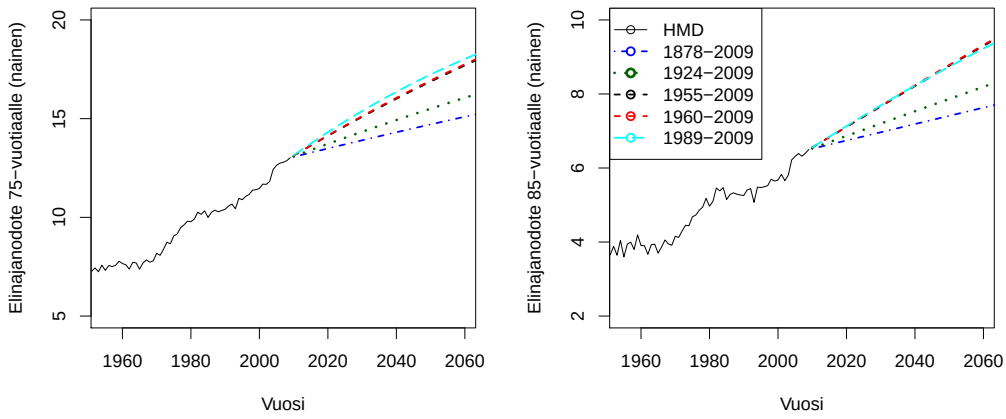



Kuva 4.8: 25- ja 50-vuotiaiden naisten elinajanodotteet eri aikaperiodeilla

Taulukosta sekä kuvista 4.8 ja 4.9 huomataan, että lähtövuodet 1960–2009 ja 1955–2009 antavat lähes samat elinajanodotteet.

Tässä raportissa aikaperiodiksi valittiin 1955–2009. Valintaan vaikutti muun muassa se, että shokkityyppistä kuolevuuden kasvua ei haluttu ottaa huomioon kuolevuusennusteessa. Toisaalta lyhimmän aikaperiodin 1989–2009 havaittiin olevan liian lyhyt pitkänajan ennusteen tekemiseen. Lähtövuosilla 1955–2009 ja 1960–2009 tehdyt kuolevuusennusteet eivät taas poikenneet toisistaan merkittävästi, joten näistä valittiin pidempi lähtöaineisto.

Samaa lähtöaineistoa 1955–2009 on käytetty TVL:n raportissa (Kuusela & Kukkala (2010)). Kyseisessä raportissa vertailtiin ennustetta myös YK:n ja Tilastokeskuksen ennusteisiin. Lisäksi vertailua tehtiin muiden maiden ennusteisiin. TVL:n tekemä Lee-Miller -ennuste ikäperiodina 1955–2008 sijoittui Tilastokeskuksen ennusteen ja YK:n ennusteen väliin. Lisäksi raportissa arvioitiin $b_x$-kertoimista, että Ruotsin ja Japanin kuolevuuden aleneminen on siirtynyt ja tulee siirtymään nuoremmista vanhempiin ikäluokkiin.




Kuva 4.9: 75- ja 85-vuotiaiden naisten elinajanodotteet eri aikaperiodeilla

## 4.4 Ennusteen epävarmuus

Ennusteessa epävarmuutta syntyy sekä parametrien että mallin epävarmuuksista. Näin ollen vuoden h ennusteen epävarmuus koostuu Lee-Miller -mallin epävarmuudesta $\sigma_{\epsilon_{x,T+h}}$ sekä parametrien $a_x$ ja $b_x$ epävarmuudesta $\sigma_{a_x}$, $\sigma_{bx}$. Parametrin $k_t$ epävarmuus jakautuu kahteen osaan: aikasarjamallin epävarmuuteen $\sigma_e$ sekä aikasarjamallin vakion $c$ epävarmuuteen $\sigma_c$. Tietyillä oletuksilla voidaan osoittaa (Lee & Carter (1992)), että virhetermi on muotoa:

$$\sigma^2_{E_{x,T+h}} = \sigma^2_{\epsilon_{x,T+h}} + \sigma^2_{a_x} + \hat{b}^2_x(h\sigma^2_e + (h\sigma_c)^2) + \sigma^2_{b_x}(\hat{k}^2_{T+h} + (h\sigma^2_e + (h\sigma_c)^2))$$

Lee ja Carter osoittivat artikkelissaan, että parametrin $k_t$ epävarmuus on dominoivaa pitkissä ennusteissa. Toisin sanoen parametreihin $a_x$ ja $b_x$ liittyvä epävarmuus on vähäistä pitkällä aikavälillä. Tässä raportissa tehdään ennuste 120 vuodelle, joten on perusteltua ottaa huomioon vain parametriin $k_t$ liittyvä epävarmuus. Näin ollen normaalijakaumaoletukselle perustuva luottamusväli piste-estimaatille luottamustasolla $\alpha$ on seuraava:

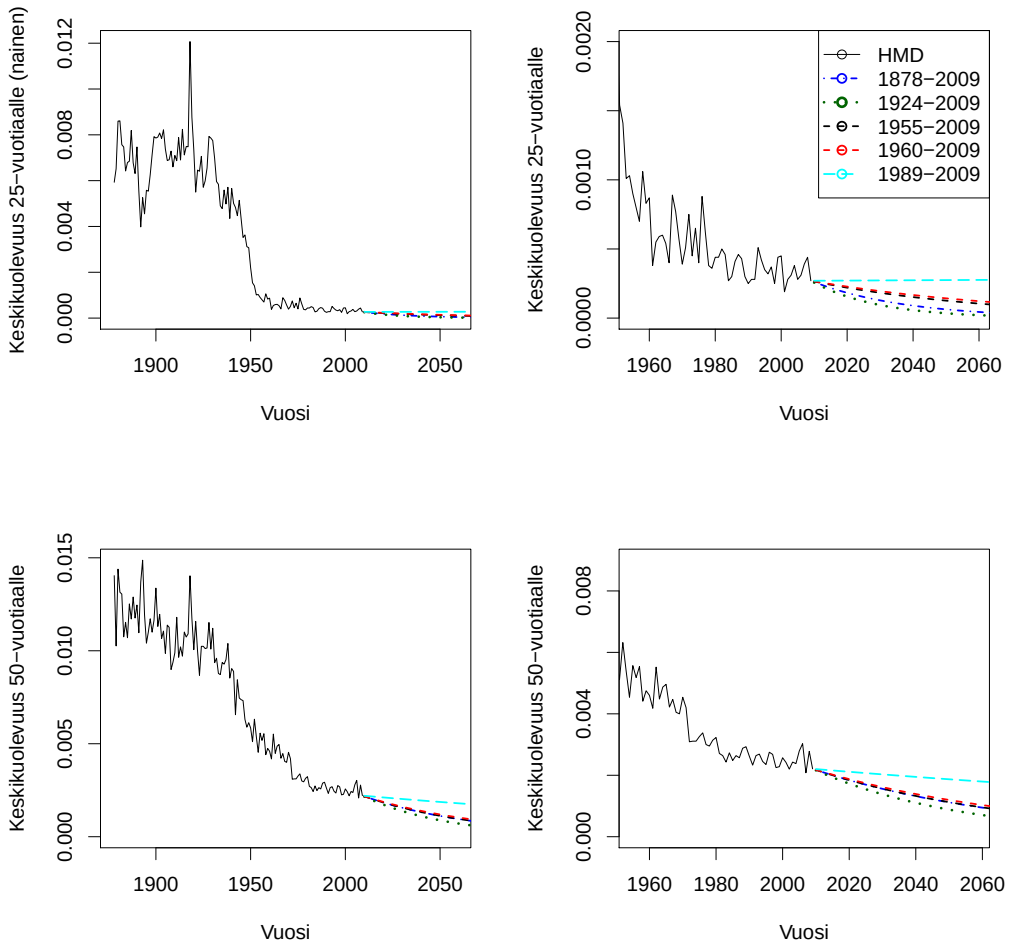
$$m_{x,T+h} exp\left(\pm z_{\alpha/2}\sqrt{b^2_x(h\sigma^2_e + (h\sigma_c)^2)}\right),$$

missä $z_{\alpha/2}$ on standardoitua normaalijakaumaa luottamustasolla $\alpha$ vastaava prosenttipiste ja piste-estimaatti on muotoa

$$\hat{m}_{x,T+h} = m_{x,T}\exp(b_x h\hat{c}).$$

Myöhemmin esiteltävät luottamusvälit raportissa ja liitteissä perustuvat edellä esitettyyn kaavaan.

Jos luottamusvälejä halutaan soveltaa periodiaineistosta muodostettuun kohorttiaineistoon, kannattaa parametri $b_x$ vakioida (PricewaterhouseCoopers (2011)). Näin hajonta riippuu ainoastaan ennusteen pituudesta. Turvaavan luottamusvälin saa käyttämällä parametrin $b_x$ lokaalia maksimia ikävälillä 20–110.



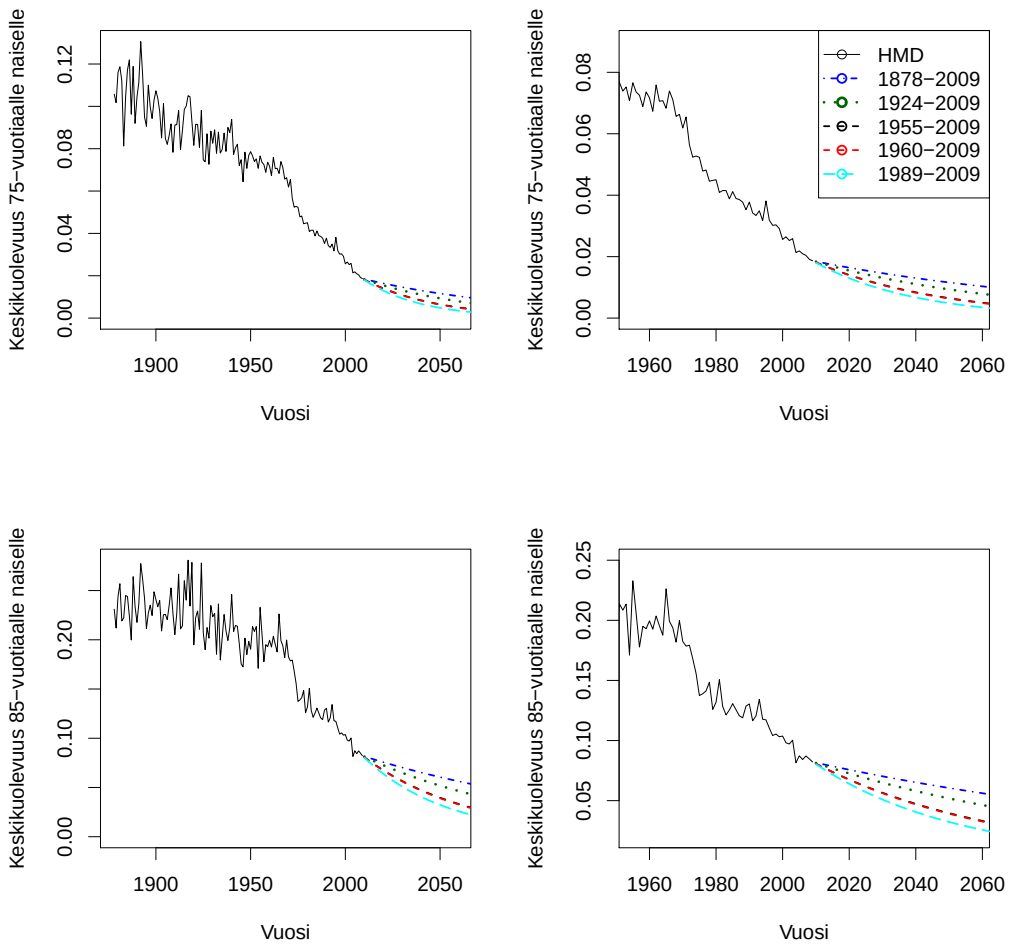

Kuva 4.10: 25- ja 50-vuotiaiden naisten keskikuolevuusennusteet eri aikaperiodeilla

Kuolevuuden epävarmuus voidaan ottaa huomioon myös simuloimalla. Tällainen MCMC- eli Markov Chain Monte Carlo -simulointialgoritmi on kuvattu Ronkaisen väitöskirjassa luvussa 4.4 s. 76 (Ronkainen (2012)).

## 4.5 Lopullinen ennuste ja vertailut aikaisempiin tutkimuksiin

Ennuste tehtiin Lee-Miller -menetelmällä käyttäen lähtöaineistoa vuosilta 1955–2009. Aineistona on käytetty HMD-tietokannan tasoitettua kuolevuusaineistoa ikävälille 0–110+ (merkintä 110+ tarkoittaa, että yli 109-vuotiaat on huomioitu ikäluokassa 110+). Liitteessä D on esitetty muun muassa ennusteen $m_{x,t}$-luvut. Kuolevuusennuste on tehty 120 vuodelle, jotta aineistosta voi muodostaa kohorttiaineiston.

Tuloksia on vertailtu TVL:n tutkimukseen, sillä se on tuorein vastaava tutkimus, jossa ennustemenetelmänä on käytetty Lee-Miller -menetelmää

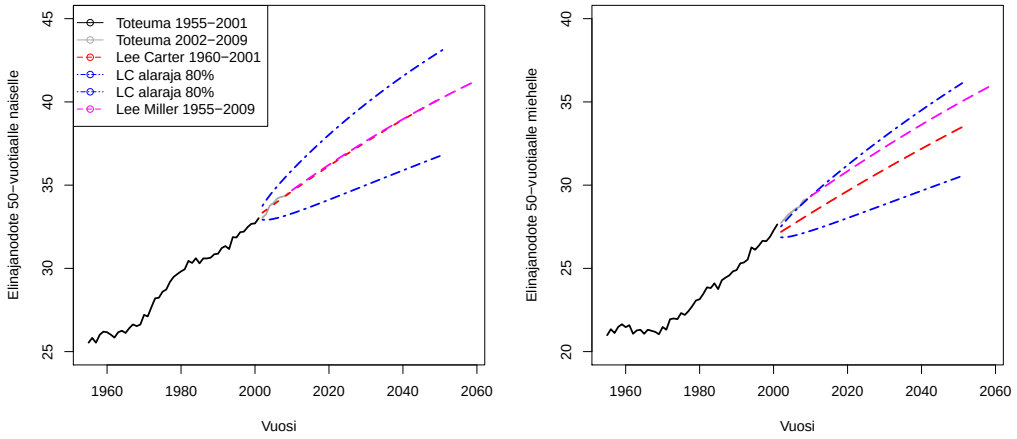



Kuva 4.11: 75- ja 85-vuotiaiden naisten keskikuolevuusennusteet eri aikaperiodeilla

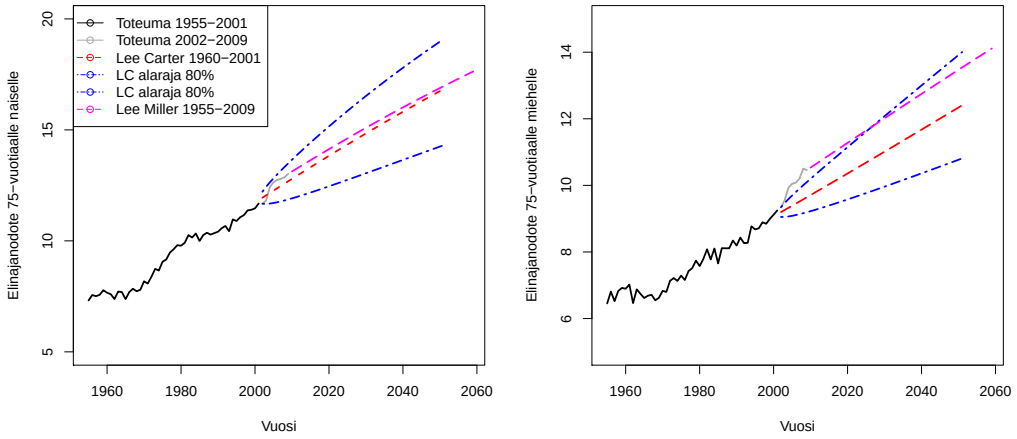
(Kuusela & Kukkala (2010)). Lisäksi myös TVL:n tutkimus on toteutettu Hyndmanin R-paketilla (versio 1.03). Liitteessä C on esitetty vertailun tulokset. Vertailussa elinajanodotteet on laskettu periodeittain. Ikäväli on tasoitetussa aineistossa 0-110+ ja tasoittamattomassa aineistossa 0–100+. Tasoittamattoman aineiston ikävälin rajoituksen johdosta erot tasoitetun ja tasoittamattoman aineiston välillä ovat pieniä.

K2004-referenssikuolevuudesta ei ollut saatavilla tietoja Lee-Carter -mallin tuloksista laisinkaan. Näin ollen vertailua K2004-referenssikuolevuustuloksiin tässä yhteydessä ei voida tehdä. Vertaillaan kuitenkin edellistä ennustetta ja uutta ennustetta toisiinsa seuraavasti: verrataan Lee-Carter -ennustetta, jossa lähtöaineistona ovat vuodet 1960–2001 (K2004-tutkimusta vastaava lähtödata), Lee-Miller -ennusteeseen, jossa lähtövuodet ovat 1955–2009. Kummatkin ennusteet on toteutettu Rob J. Hyndmanin R-paketilla. Kuvissa 4.12 ja 4.13 on esitetty ennusteet elinajanodotteille. Kuvista huomataan aikaisemmassa ennusteessa syntynyt Lee-Carter -menetelmälle ominainen ”jump-off”-virhe. Kuvasta 4.13 huomataan, että 75-vuotiaalla miehellä havaitut kuolevuudet vuosilta 2002–2008 eivät sovi Lee-Carter -ennusteen luottamusvä-

lin sisäpuolelle. Esitetyissä tapauksissa useimmiten miesten uusi Lee-Miller -ennuste lähtövuosilla 1955–2009 kulkee edellisen Lee-Carter -ennusteen lähtövuosina 1960–2001 tuottaman 80 %:n luottamusvälin ylärajan tuntumassa. Naisten vastaava Lee-Miller -ennuste kulkee selkeästi Lee-Carter -ennusteen luottamusvälin sisäpuolella.




Kuva 4.12: 50-vuotiaiden naisten ja miesten elinajanodotteet Lee-Carter -menetelmällä ja 80 %:n luottamusvälit (lähtöaineisto 1960–2001, kuvassa lyhennetty ”LC”). Lisäksi Lee-Miller -ennuste lähtöaineistolla 1955–2009.




Kuva 4.13: 75-vuotiaiden naisten ja miesten elinajanodotteet Lee-Carter -menetelmällä ja 80 %:n luottamusvälit (lähtöaineisto 1960–2001, kuvassa lyhennetty ”LC”). Lisäksi Lee-Miller -ennuste lähtöaineistolla 1955–2009.

Seuraavissa taulukoissa 4.1 ja 4.2 on esitetty vastaavat elinajanodotteet vuosille 2010, 2030 ja 2050 Lee-Carter- ja Lee-Miller -menetelmillä. Lee-Carter -menetelmän tulokset on esitetty eri aikaperiodeilla. Huomataan, että miehillä elinajanodotteet ovat odotetusti kasvaneet viime ennusteesta. Naisilla taas muutos on ollut pieni.

Taulukko 4.1: Lähtövuosien lisäyksen vaikutus elinajanodotteisiin (miehet)

| Miehet | | Lee-Carter 1955–2000 | 1960–2000 | 1955–2001 | 1960–2001 | Lee-Miller 1955–2009 |
|---|---|---|---|---|---|---|
| 0-vuotias | 2010 | 75.6 | 75.8 | 75.5 | 75.6 | 76.7 |
| | 2030 | 78.8 | 79.1 | 78.6 | 78.8 | 80.3 |
| | 2050 | 81.7 | 82.0 | 81.4 | 81.7 | 83.4 |
| 25-vuotias | 2010 | 51.3 | 51.5 | 51.2 | 51.3 | 52.4 |
| | 2030 | 54.3 | 54.6 | 54.1 | 54.4 | 55.8 |
| | 2050 | 57.1 | 57.4 | 56.8 | 57.1 | 58.8 |
| 50-vuotias | 2010 | 28.2 | 28.5 | 28.1 | 28.3 | 29.3 |
| | 2030 | 30.9 | 31.1 | 30.7 | 30.9 | 32.3 |
| | 2050 | 33.3 | 33.6 | 33.1 | 33.4 | 34.9 |
| 75-vuotias | 2010 | 9.7 | 9.8 | 9.7 | 9.7 | 10.5 |
| | 2030 | 11.1 | 11.2 | 10.9 | 11.0 | 12.0 |
| | 2050 | 12.4 | 12.5 | 12.3 | 12.3 | 13.5 |

Taulukko 4.2: Lähtövuosien lisäyksen vaikutus elinajanodotteisiin (naiset)

| Naiset | | Lee-Carter 1955–2000 | 1960–2000 | 1955–2001 | 1960–2001 | Lee-Miller 1955–2009 |
|---|---|---|---|---|---|---|
| 0-vuotias | 2010 | 82.6 | 82.8 | 83.3 | 83.4 | 83.3 |
| | 2030 | 85.9 | 85.9 | 86.7 | 86.7 | 86.7 |
| | 2050 | 88.7 | 88.7 | 89.5 | 89.5 | 89.5 |
| 25-vuotias | 2010 | 58.1 | 58.2 | 58.7 | 58.8 | 58.8 |
| | 2030 | 61.1 | 61.2 | 61.9 | 61.9 | 62.0 |
| | 2050 | 63.9 | 63.9 | 64.6 | 64.7 | 64.7 |
| 50-vuotias | 2010 | 33.9 | 34.0 | 34.4 | 34.6 | 34.7 |
| | 2030 | 36.7 | 36.9 | 37.4 | 37.6 | 37.6 |
| | 2050 | 39.3 | 39.4 | 40.0 | 40.2 | 40.2 |
| 75-vuotias | 2010 | 12.3 | 12.4 | 12.6 | 12.8 | 13.1 |
| | 2030 | 14.2 | 14.3 | 14.7 | 14.8 | 15.1 |
| | 2050 | 16.1 | 16.2 | 16.6 | 16.7 | 16.9 |

Mika Mäkisen SHV-työssä esitetyt 0-vuotiaan elinajanodotteet on esitetty taulukossa 4.3 (Mäkinen (2004)). Mäkisen SHV-työssä Lee-Carter -malli tehtiin ikävälille 0-110+. Lähtövuosina oli käytetty vuosia 1955–2000. Elinajanodotteet ovat samaa luokkaa taulukoissa 4.1 ja 4.2 esitettyjen elinajanodotteiden kanssa. Pienet erot selittyvät elinajanodotteen laskentatavasta.

Taulukko 4.3: Mika Mäkisen SHV-työssä esitetyt 0-vuotiaan elinajanodotteet

| | Naiset | Miehet |
|---|---|---|
| 2030 | 85.94 | 78.86 |
| 2050 | 88.79 | 81.74 |
| 2100 | 94.28 | 87.82 |

# 5 Henkivakuutuksen korjauskertoimet

## 5.1 Yleistä

Luvun 4 ennusteita ei sellaisenaan voi soveltaa referenssiperusteessa, koska henkivakuutetuilla kuolevuus on alhaisempaa kuin Suomen väestön keskimääräinen kuolevuus. Tämä johtuu muun muassa vakuutusyhtiöiden tekemästä vastuunvalinnasta. Toisaalta henkivakuutusaineisto ei ole riittävän kattava pitkän ajan ennusteen tekemiseen. Näin ollen tässä luvussa henkivakuutusaineiston ja Suomen väestöaineiston avulla muodostetaan korjauskerroin, jonka avulla voidaan korjata Suomen väestön kuolevuusennusteita henkivakuutettujen kuolevuutta paremmin vastaavaksi.

## 5.2 Lukumäärä- ja summakuolevuuksien vertailu

Aikaisemmissa henkivakuutuksen kuolevuustutkimuksissa on käytetty summakuolevuutta. Summakuolevuuden etuina on pidetty muun muassa sen kykyä ottaa huomioon henkilön varallisuus, mikä korreloi henkilön sosiaaliseen asemaan. Henkivakuutuksen nykyisessä referenssikuolevuudessa K2004 sekä TyEL:n kuolevuusperusteessa on käytetty volyymisuureita eli riskisummia ja rahastoja. Toisaalta viimeisimmissä lakisääteisen tapaturmavakuutuksen ja liikennevakuutuksen kuolevuusreferenssimalleissa vakuutusaineistona käytettiin lukumääräaineistoa (Kauppi (2004), PricewaterhouseCoopers (2011)).

Seuraavaksi vertaillaan summa- ja lukumääräkuolevuuksien eroja henkivakuutusaineistosta tuotteittain. Tämän raportin yhteydessä kerätty volyymisuureaineisto on laadultaan lukumääräaineistoa heikompi ja osin puutteellinen. Tämän takia volyymisuureaineistosta on jouduttu poistamaan kokonaan yhden suuren yhtiön tiedot. Vertailuja ei voida myöskään tehdä tapaturmaisen kuoleman osalta aineiston puutteiden vuoksi. Yksilöllisten säästö- ja eläkevakuutusten sijoitussidonnaiset vakuutukset on jätetty pois tarkasteluista, koska niissä säästöjen määrä vaihtelee taustalla olevan sijoitusportfolion volatiliteetin mukaan ja tyypillisesti summat ovat pieniä.

Lukumääräkuolevuus $m_x^L$ on kuolleiden lukumäärän ($D_x$) suhde altistumaan ($E_x$) (jatkossa "D/E"-luku) tarkastelujaksolla

$$m_x^L = \frac{\sum_t D_{x,t}}{\sum_t E_{x,t}}.$$

Summakuolevuus $m_x^S$ määritellään tässä raportissa seuraavasti:

$$m_x^S = \frac{\sum_t S_{x,t}^K}{\sum_t \left(S_{x,t} - \frac{1}{2} S_{x,t}^K\right)},$$

missä $S_{x,t}^K$ on $x$-ikäisenä kuolleiden volyymisuure vuonna $t$ ja $S_{x,t}$ $x$-ikäisten volyymisuure vuoden $t$ alussa. Aineiston negatiiviset volyymisuureet on muutettu tarkasteluissa nolliksi. Raportissa esitetyissä kuvaajissa 5.1–5.5 kuolevuudet on laskettu ikäluokittain. Summaus on tehty yli vuosien (2002–2010 ja ryhmäeläkkeen osalta 2002–2009) satunnaisvaihtelun vähentämiseksi. Lisäksi kuvissa on esitetty summa- ja lukumääräkuolevuuksien suhde väestökuolevuuteen ($m_{x,t}$), missä on käytetty HMD-tietokannan väestökuolevuuksien keskiarvoa vuosilta 2002–2009 (jatkossa $\frac{m_x^S}{m_x}$).

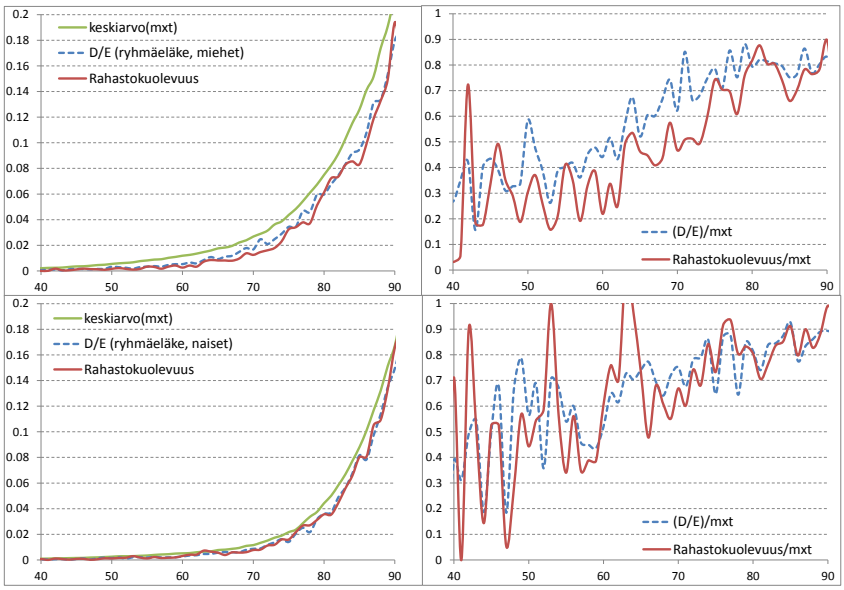
Aikaisemmin henkivakuutusten riskiperusteanalyyseissä on käytetty volyymisuureena niin kutsuttua riskisummaa. Analyysien ohjeissa se on määritelty kuoleman varalta olevan summan ja rahaston erotukseksi (Seppo Koskinen ja Hannu Vaajama, SHY (1982)). Yhtiöiden käytänteissä on kuitenkin ollut eroja. Koska vastaavasta kuolinmenosta ei ole vähennetty vapautuvaa rahastoa, eräät yhtiöt ovat riskisummaksi määritelleet kuoleman varalta olevan riskisumman. Tässä tutkimuksessa volyymisuureena on käytetty riskiperusteanalyysin mukaista riskisummaa. Eläkevakuutuksen volyymisuureena on käytetty rahastoa.

Perinteisessä ryhmäeläkevakuutuksessa rahasto määräytyy vakuutetun palkan mukaan, joten summa- eli tässä tapauksessa rahastokuolevuudessa ansiotason ja kuolevuuden negatiivinen korrelaatio tulee otettua huomioon. Kuvassa 5.1 on esitetty ryhmäeläkkeen summa- sekä lukumääräkuolevuudet miehille ja naisille. Huomataan, että miehillä ero näiden kuolevuuksien välillä on selvä. Naisilla summa- ja lukumääräkuolevuudet ovat lähellä toisiaan. Toisin sanoen miehillä ansiotaso vaikuttaa kuolevuuteen enemmän kuin naisilla. Miesten keskimääräiset volyymisuureet ovat korkeampia kuin naisilla.

Kuvissa 5.2 ja 5.3 on esitetty summa- ja lukumääräkuolevuudet yksilölliselle riskivakuutukselle sekä ryhmäriski- ja ryhmäetuvakuutuksille. Huomataan, että miehillä summakuolevuus on lukumääräkuolevuutta alempaa, kun taas naisilla kuolevuuksien välillä ei ole merkittävää eroa, kun otetaan huomioon korjauskertoimeen liittyvä epävarmuus.

Yksilöllisen säästövakuutuksen tapauksessa summa- ja lukumääräkuolevuudet ovat lähes samat niin miehillä kuin naisilla, varsinkin kun otetaan huomioon suureisiin liittyvä epävarmuus (kuva 5.4). Yksilöllisessä eläkevakuutuksessa eroa lukumäärä- ja summa- eli tässä tapauksessa rahastokuolevuuksien välillä on noin yli 50-vuotiailla (kuva 5.5). Miehillä ero on suurempaa kuin naisilla. Vastaava ilmiö todettiin miesten osalta ryhmäeläkkeissä, joskin suurempana.

Kuvien 5.1–5.5 perusteella miehillä summa- ja lukumääräkuolevuuksien ero on suurempaa kuin naisilla. Tähän selityksenä saattaa olla erot miesten ja naisten volyymisuureiden jakaumissa. Summa- ja lukumääräkuolevuuksien tuotteittaiset erot selittyvät muun muassa volyymisuureiden kyvyllä kuvata varallisuutta. Esimerkiksi lainaturvavakuutuksissa volyymisuure, joka on tässä määritelty riskisummaksi, pienenee vakuutussopimuksen aikana vaikka



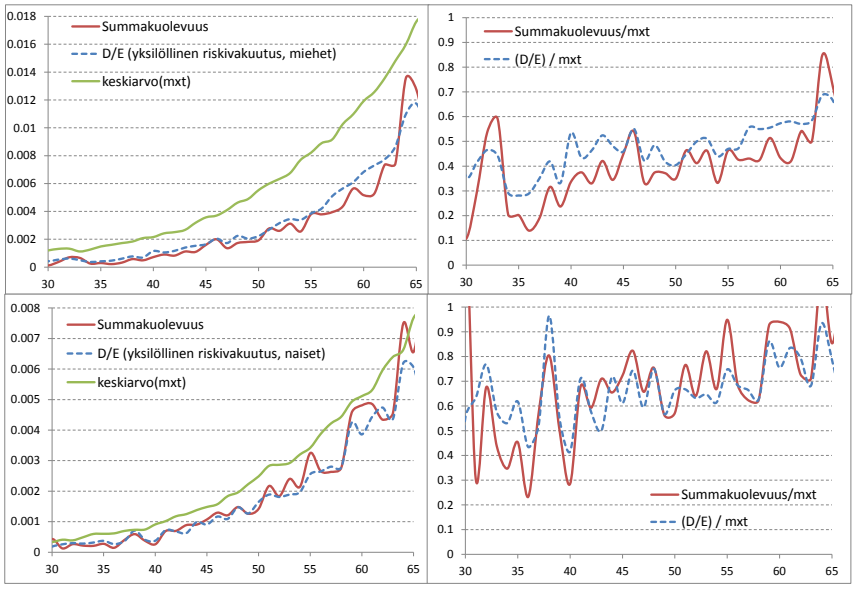

Kuva 5.1: Ryhmäeläkkeiden lukumäärä- ja rahastokuolevuusvertailu miehille ja naisille. Kuvassa väestökuolevuus $m_{x,t}$ on keskiarvo vuosilta 2002–2009 (HMD).

ansiotaso ei muuttuisi. Tässä tapauksessa riskisumma kuvaa vakuutusyhtiön vastuuta mutta se ei välttämättä kuvaa vakuutetun ansiotasoa. Tämä selittää sitä, että summa- ja lukumääräkuolevuudet ovat lähellä toisiaan. Kaikissa vakuutuslajeissa volyymisuureen merkitys verrattuna lukumääräkuolevuuteen on pienempi kuin lukumääräkuolevuuden merkitys väestökuolevuuteen.

Suurimmat erot summa- ja lukumääräkuolevuuksien välillä olivat miesten ryhmäeläkevakuutuksissa. Toisaalta ryhmäeläkevakuutuksia ei ollut aikaisempien tutkimusten tutkimusaineistossa. Näin ollen myös K2004-kuolevuudessa lukumääräkuolevuus lienee ollut varsin lähellä summakuolevuutta.

Vakuutusyhtiöissä myös vakuutuskannan koko ja rakenne vaikuttavat lukumäärä- ja summakuolevuuksien eroihin. Tässä raportissa tavoitteena on määrittää tuotteiden keskimääräinen kuolevuus. Kun otetaan huomioon kuolevuusennusteeseen ja korjauskertoimeen liittyvä epävarmuus, pienet erot lukumäärä- ja summakuolevuuksien välillä eivät vaikuta merkittävästi valittavaan lopulliseen referenssikuolevuuteen.

Koska lukumääräkuolevuus on luvussa 4 esitetyn väestökuolevuusennusteen pohjana, myös korjauskertoimet on mielekästä määritellä lukumääräaineiston pohjalta kaikille henkivakuutustuotteille. Jos tässä tutkimuksessa olisi valittu lähtökohdaksi summakuolevuus, lähtöaineistosta olisi jouduttu jättämään aineiston puutteiden takia huomattava osa pois. Kun lisäksi lukumääräkuolevuuden käyttö on osoittautunut riittävän tarkaksi lähes kaikissa tapauksissa, tässä tutkimuksessa keskitytään lukumääräkuolevuuteen. Kuitenkin miesten ryhmäeläkevakuutuksen osalta tullaan esittämään korjaus-




Kuva 5.2: Yksilöllisen riskivakuutusten lukumäärä- ja summakuolevuusvertailu miehille ja naisille. Kuvassa väestökuolevuus $m_{x,t}$ on keskiarvo vuosilta 2002–2009 (HMD).

kerroin myös summakuolevuuteen perustuen lineaarisella regressioanalyysillä, koska summakuolevuus on selvästi lukumääräkuolevuutta pienempi. Summakuolevuutta käytettäessä oletuksena on, että oman vakuutuskannan rakenne vastaa alan vakuutuskannan rakennetta. Regressioanalyysi on sovitettu suureeseen summakuolevuuden suhde väestökuolevuuteen $\frac{m_x^S}{m_x}$.

## 5.3 Poisson-jakauma

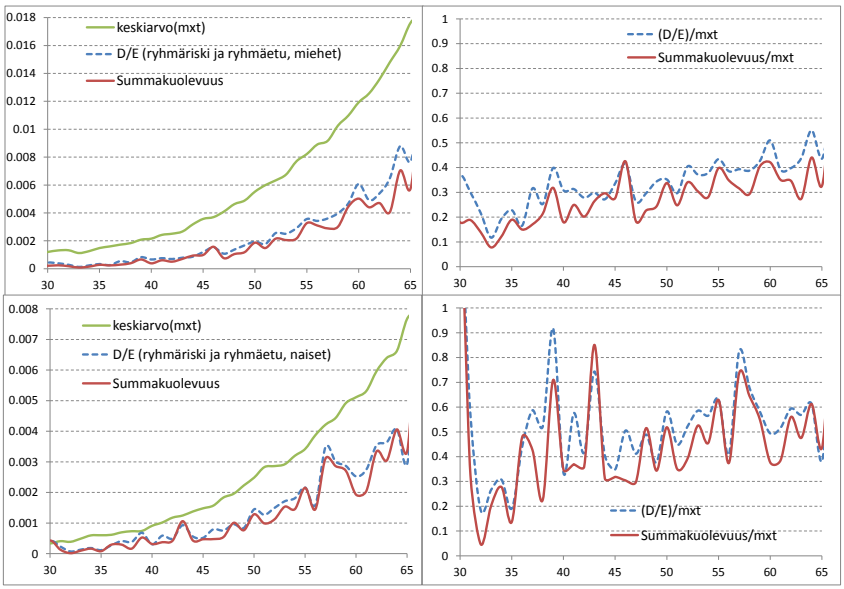
Poisson-jakauma on diskreetti todennäköisyysjakauma, jonka arvojoukko on luonnollisten lukujen joukko. Se ilmaisee todennäköisyyden tapahtumien lukumäärälle kiinteällä aikavälillä. Oletetaan, että satunnaismuuttuja $Y$ noudattaa Poisson-jakaumaa parametrilla $\mu$, $\mu > 0$ eli $Y \sim Poi(\mu)$. Tällöin

$$P(y = j) = \frac{\mu^j}{j!} e^{-\mu}, j = 0, 1, 2, ... \tag{5.1}$$

Voidaan osoittaa, että $\mathrm{E}(y) = \sum_{j=0}^{\infty} jP(y = j) = \mu = \mathrm{Var}(y)$. Poisson-jakaumalle tyypillinen ominaisuus on muun muassa sen ei-negatiivisuus.

## 5.4 Poisson-regressio

Poisson-regressio kuuluu yleistettyjen lineaaristen mallien (”generalized linear models”) joukkoon. Se soveltuu erityisesti melko suurille aineistoille, joissa




Kuva 5.3: Ryhmäriski- ja ryhmäetuvakuutusten lukumäärä- ja summakuolevuusvertailu miehille ja naisille. Kuvassa väestökuolevuus $m_{x,t}$ on keskiarvo vuosilta 2002–2009 (HMD).

yksittäisen tapahtuman todennäköisyys on pieni. Peruslähtökohtana Poisson-regressiossa on, että tapahtumien määrä noudattaa Poisson-jakaumaa, jonka odotusarvo on $\mu$. Poisson-regressiossa odotusarvon annetaan kuitenkin vaihdella sen mukaan, kuinka monta henkilöä tai henkilövuotta $E_i$ (”exposure”=altistuminen) on ollut alttiina riskille tarkasteluaikana $i$ siten, että

$$\mathrm{E}(y_i) = E_i\mu.$$

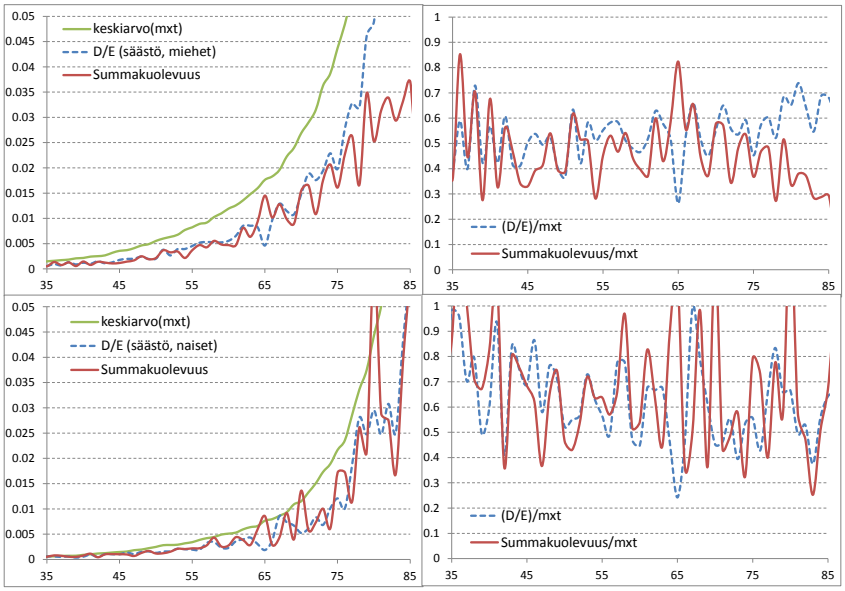
Näin ollen myös varianssin oletetaan vaihtelevan odotusarvon mukaisesti. Yleisemmin ottaen odotusarvon annetaan vaihdella sopivasti valittujen selittävien muuttujien mukaisesti. Poisson-regression tapauksessa on luontevaa määritellä, että

$$\mathrm{E}(y_i) = \mu_i = \exp(\mathbf{z_i'}\boldsymbol{\beta}),$$

missä $\mathbf{z_i} = 1, z_{i1}, ..., z_{ip}$ on havaintoon $i$ liittyvä selittävien muuttujien vektori ja kertoimet $\beta_j$ $(j = 0, ..., p)$ ovat regressiokertoimia. Kertoimet $\beta_j$ voidaan estimoida suurimman uskottavuuden menetelmällä (Alho & Spencer (2005), Dobson (2002), Kaas et al. (2008)).

Eksponenttifunktion käyttö systemaattisen osan $\mathbf{z_i'}\boldsymbol{\beta}$ määrittelyssä varmistaa, että odotusarvo $\mathrm{E}(y_i)$ pysyy positiivisena. Funktio, joka yhdistää odotusarvon systemaattiseen osaan, niin sanottu logaritminen linkkifunktio, on muotoa $\ln(\mu_i) = \mathbf{z_i'}\boldsymbol{\beta}$. $\ln(\mu_i)$ on siis selittävien muuttujien lineaarinen funktio.

Mallissa voidaan valita niin sanottuja ”offset”-muuttujia, joiden regressiokerroin pakotetaan ykköseksi. Toisin sanoen oletetaan, että näiden muuttu-




Kuva 5.4: Yksilöllisen säästövakuutuksen lukumäärä- ja summakuolevuusvertailu miehille ja naisille (perustekorkoiset tuotteet). Kuvassa väestökuolevuus $m_{x,t}$ on keskiarvo vuosilta 2002–2009 (HMD).

jien arvot ovat tunnettuja. Esimerkiksi muuttuja $E_i$ asetetaan usein "offset"-muuttujaksi. Tällöin linkkifunktio on seuraava:

$$\ln(\mu_i) = \ln(E_i) + \mathbf{z_i'}\boldsymbol{\beta}$$

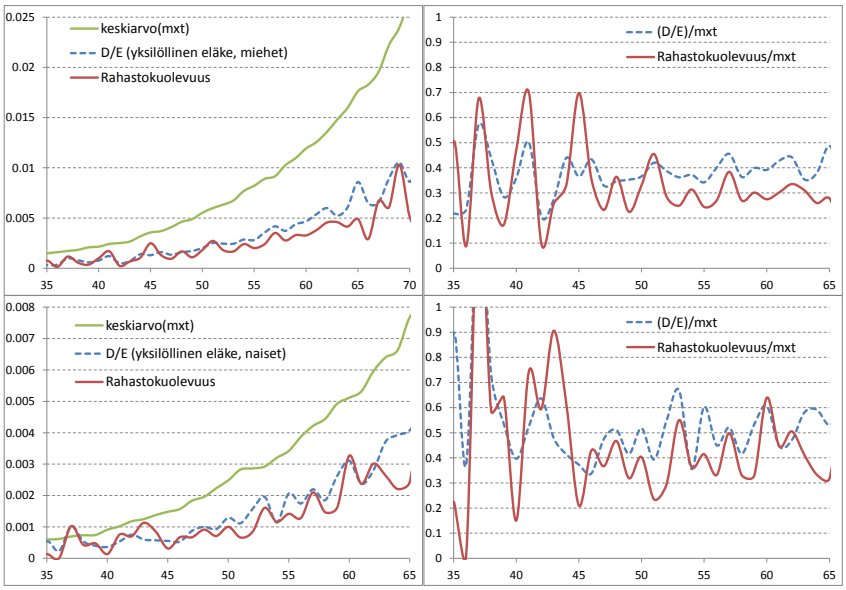
## 5.5 Korjauskertoimen määrittäminen

Henkivakuutusyhtiöiden referenssikuolevuudessa (K2004) henkivakuutuskuolevuuksien ja populaatiokuolevuuksien suhdetta arvioitiin riskiperusteanalyysistä (SHY (1982)) laskettujen viisivuotisluokittaisten kuolevuuksien sekä Suomen väestön vastaavien kuolevuuslukujen suhteen avulla seuraavasti ($sp = 0$ miehillä ja $sp = 1$ naisilla):

$$p_x^{sp} = \frac{\mu_x^{\text{henkiaineisto}}}{\mu_x^{\text{populaatio}}}$$

Tarkastelussa oletettiin, että henkivakuutuskuolevuus vastaa väestökuolevuutta ikävuosina 0–20 sekä yli 70-vuotiailla. Lisäksi suhteen oletettiin pysyvän samana tulevaisuudessa. Kuvassa 5.6 on esitetty Mäkisen SHV-työssä esitetyt korjauskertoimet. Naisilla korjauskertoimet ovat korkeammalla tasolla kuin miehillä.

Tässä raportissa samaa ongelmaa lähestytään lukumääräaineiston avulla. Jotta kertoimien määrittämisessä tarvittaisiin vähemmän asiantuntija-




Kuva 5.5: Yksilöllisen eläkevakuutuksen lukumäärä- ja rahastokuolevuusvertailu miehille ja naisille (perustekorkoiset tuotteet). Kuvassa väestökuolevuus $m_{x,t}$ on keskiarvo vuosilta 2002–2009 (HMD).

harkintaa, menetelmäksi valittiin Poisson-regressio, jossa estimointi tapahtuu suurimman uskottavuuden menetelmällä (luku 5.4). Vastaavaa menetelmää käytettiin lakisääteisen tapaturmavakuutuksen ja liikennevakuutuksen kuolevuusreferenssimallissa (PricewaterhouseCoopers (2011)).

Mallissa oletetaan, että kuolleiden lukumäärä $D_x$ iässä $x$ noudattaa Poisson-jakaumaa odotusarvonaan $L_x M_x E_x$, missä iässä $x$ $E_x$=henkilövuosien lukumäärä tarkastelujaksolla, $M_x$=väestökuolevuus tarkastelujaksolla ja $L_x$= korjauskerroin väestö- ja vakuutuskuolevuuksien välillä. Näin ollen odotusarvo muodostuu korjauskertoimen ja väestökuolevuuteen perustuvien odotettujen kuolleiden lukumäärien tulona.

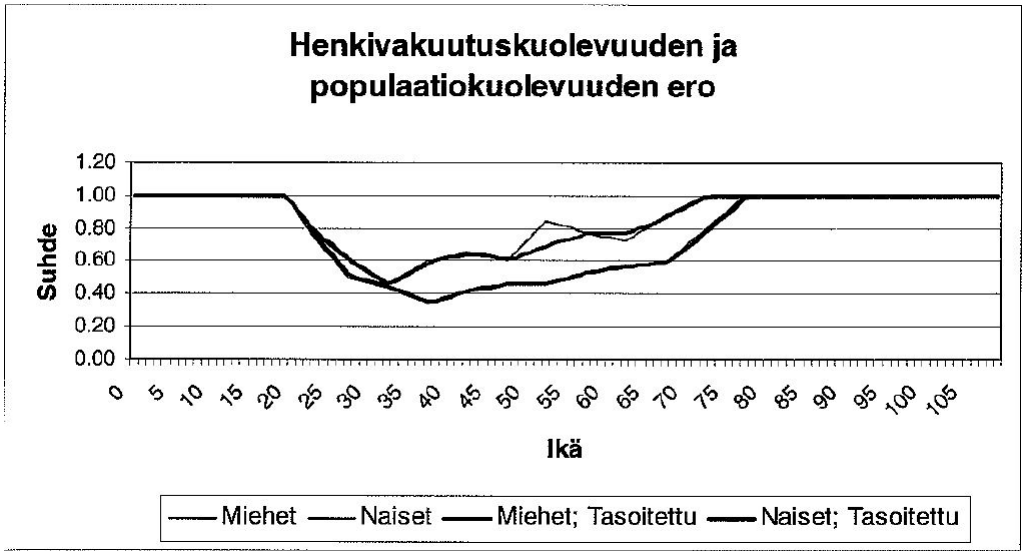
Väestökuolevuuden perusteella odotettu kuolleiden lukumäärä ($M_x E_x$) oletetaan tunnetuksi eli se asetetaan niin sanotuksi ”offset”-muuttujaksi. Näin ollen linkkifunktio on muotoa

$$\ln(D_x) = \ln(E_x) + \ln(M_x) + \mathbf{z_i}'\boldsymbol{\beta},$$

missä $\hat{\boldsymbol{\beta}}$ on parametrivektori ja $\mathbf{z_i}$ on havaintoon $i$ liittyvä selittävien muuttujien vektori. Toisin sanoen loglineaarinen malli on

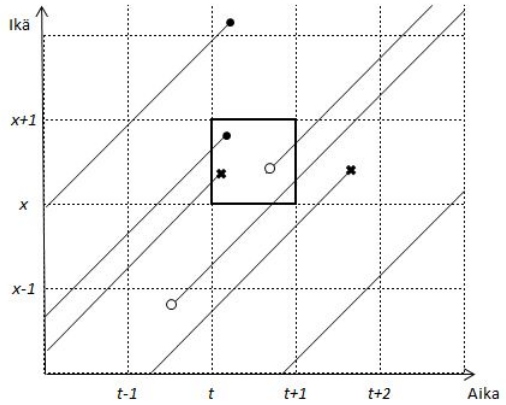
$$L_x = \exp(\mathbf{z_i}'\boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 z_{i1} + ... + \beta_p z_{ip}). \tag{5.2}$$

Aluksi Poisson-regressiota sovelletaan koko vakuutusaineistoon vuosilta $t$=2002–2009 ilman tuotteittaisia jaotteluita (luku 3.2). Vuosi 2010 jätetään pois tarkasteluista, sillä HMD-tietokannan väestökuolevuudet ovat saatavilla vain vuoteen 2009 asti. Analyysissä tarkastelujaksoksi valitaan yksi vuosi.




Kuva 5.6: K2004-korjauskertoimet

Näin ollen jokaiselle iälle $x$ lasketaan vuosittain $E_{x,t}$-luvut, jotka kertovat kuolemalle altistuneiden henkilövuosien lukumäärän iässä $x$ vuonna $t$. Näin ollen $E_{x,t}$ kertoo vakuutettujen määrän iässä $x$ tiettynä vuotena $t$. Vakuutettu otetaan huomioon, jos vakuutus on ollut koko vuoden voimassa, jos vakuutus on alkanut tai päättynyt kesken vuoden tai jos henkilö on kuollut vuoden aikana. Tätä on havainnollistettu kuvassa 5.7, jossa musta piste kuvaa vakuutuksen päättymistä ja valkoinen piste vakuutuksen alkamista. Musta rasti taas kuvaa kuolemaa. Vuoden $t$ exposure-luku $E_{x,t}$ eli altistuminen $x$-ikäiselle on siis tummennetun laatikon sisällä olevien viivojen pituus. Kuolleiden lukumäärä on mustien rastien lukumäärä tummennetussa laatikossa. Tasoittamattomat keskimääräiset väestökuolevuusluvut perustuvat HMD-tietokannan tietoihin.




Kuva 5.7: Altistumisen ($E_{x,t}$) laskennan havainnollistaminen Lexis-diagrammin avulla

## 5.6 Soveltaminen henkivakuutusaineistoon

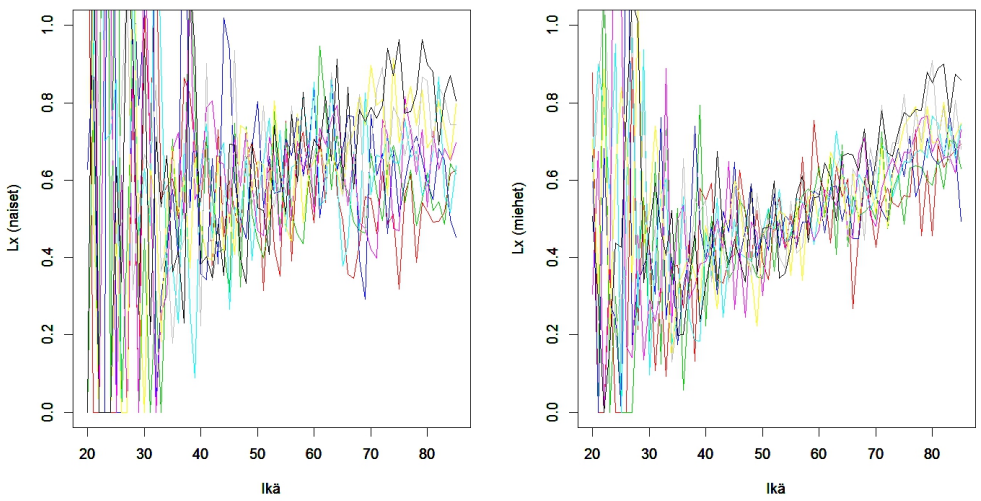
### 5.6.1 Yleistä

Kuvassa 5.8 on havainnollistettu korjauskerrointa $L_{x,t}$ henkivakuutusyhtiöiden koko aineistolle ilman vakuutuslajijaottelua. Siinä eri värit kuvaavat eri tarkasteluvuosia. Huomataan, että vaikka aineistoa ei ole ryhmitelty pienempiin tuoteryhmiin, vaihtelu on suurta erityisesti nuorissa ja vanhoissa ikäluokissa. Tämä selittyy havaintojen ja kuolintapauksien vähäisyydellä (vrt. luku 3.3). Mallia sovitettaessa sekä ryhmittelyjä tehtäessä tuleekin ottaa huomioon aineiston rajoitteet. Jos aineiston vähäisyyden vuoksi epäillään aineiston luotettavuutta, mallin luotettavuutta voidaan parantaa esimerkiksi ikien rajoittamisella tai ikäluokkatarkastelulla (esimerkiksi viisivuotisluokat).

Tämän raportin lopullisessa mallissa tehdään oletus, että ikäkohtaiset korjauskertoimet eivät riipu ajasta $t$. Näin ollen mallissa vuosimuuttuja ei ole selittävänä muuttujana. Poisson-regressio sovitetaan aineistoon, jossa korjauskerroin $L_x$ on laskettu vuosien yli summatusta aineistosta.

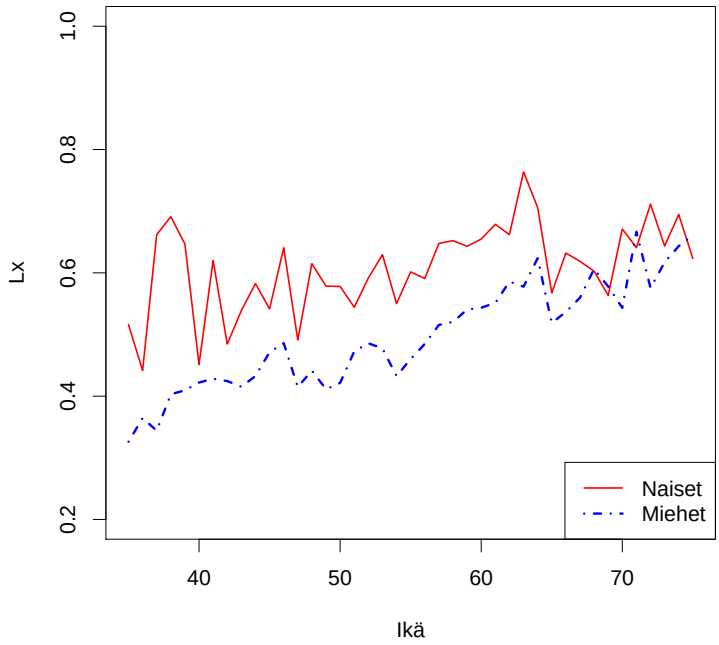
$$L_x = \frac{\sum_{t=2002}^{2009} D_{x,t}}{\sum_{t=2002}^{2009} m_{x,t} E_{x,t}}$$

Tätä oletusta puoltaa myös aineiston vähäisyys ja siitä johtuva epävarmuus vuositason aineistossa. Kuvioita 5.8 ja 5.9 verratessa huomataan summauksen tasoittava vaikutus.




Kuva 5.8: Miesten ja naisten $L_{x,t}$-suureet ilman vakuutuslajijaottelua havaintovuosittain

Kuvasta 5.9 huomataan, että miesten ja naisten korjauskertoimien $L_x$ välillä on eroja. Miehillä muun muassa ikäriippuvuus vaikuttaa olevan suurempaa kuin naisilla. Näin ollen mallissa on hyvä tutkia sukupuolimuuttujan merkitsevyyttä selittävänä muuttujana. Kuvaajissa voidaan havaita myös




Kuva 5.9: Miesten ja naisten $L_x$-suureet ilman vakuutuslajijaottelua

korjauskertoimen tason muutos eläkeiässä. Tämä johtuu ikärakenteesta sekä antaa viitteitä siitä, että korjauskerroin on erilainen eri tuotteilla.

Kuvassa 5.10 on esitetty vastaava kuva vakuutuslajeittain. Huomataan, että korjauskertoimet vaihtelevat vakuutuslajeittain. Tämä on odotettuakin, sillä eri vakuutuslajeissa muun muassa vastuunvalinta on erilaista. Säästö- ja sijoitusvakuutuksissa sekä vanhuuseläkevakuutuksissa ei välttämättä tehdä juuri lainkaan vastuunvalintaa, ellei kyse ole merkittävistä riskisummista. Toisaalta myös ryhmävakuutuksissa, joissa samalla sopimuksella vakuutetaan suuri ryhmä henkilöitä, ei ole välttämättä vakuutettutason vastuunvalintaa. Kuitenkin yksilöllisissä riskivakuutuksissa vastuunvalinnalla on tärkeä rooli. Nämä asiat perustelevat tekemään tarkasteluita erilaisilla vakuutuslajijaotteluilla, kun taas aineiston rajoitteiden takia olisi hyvä käyttää mahdollisimman vähän ryhmittelyjä. Koska tuotteittainen tarkastelu on mielekästä ja perusteltua, myöhemmin esitetään tulokset sekä koko aineistolle että erilaisilla vakuutuslajijaotteluilla. Jaotteluissa tavoitteena on löytää kuolevuudeltaan samanlaisia ryhmiä siten, että aineiston rajoitteet on parhaalla mahdollisella tavalla otettu huomioon.

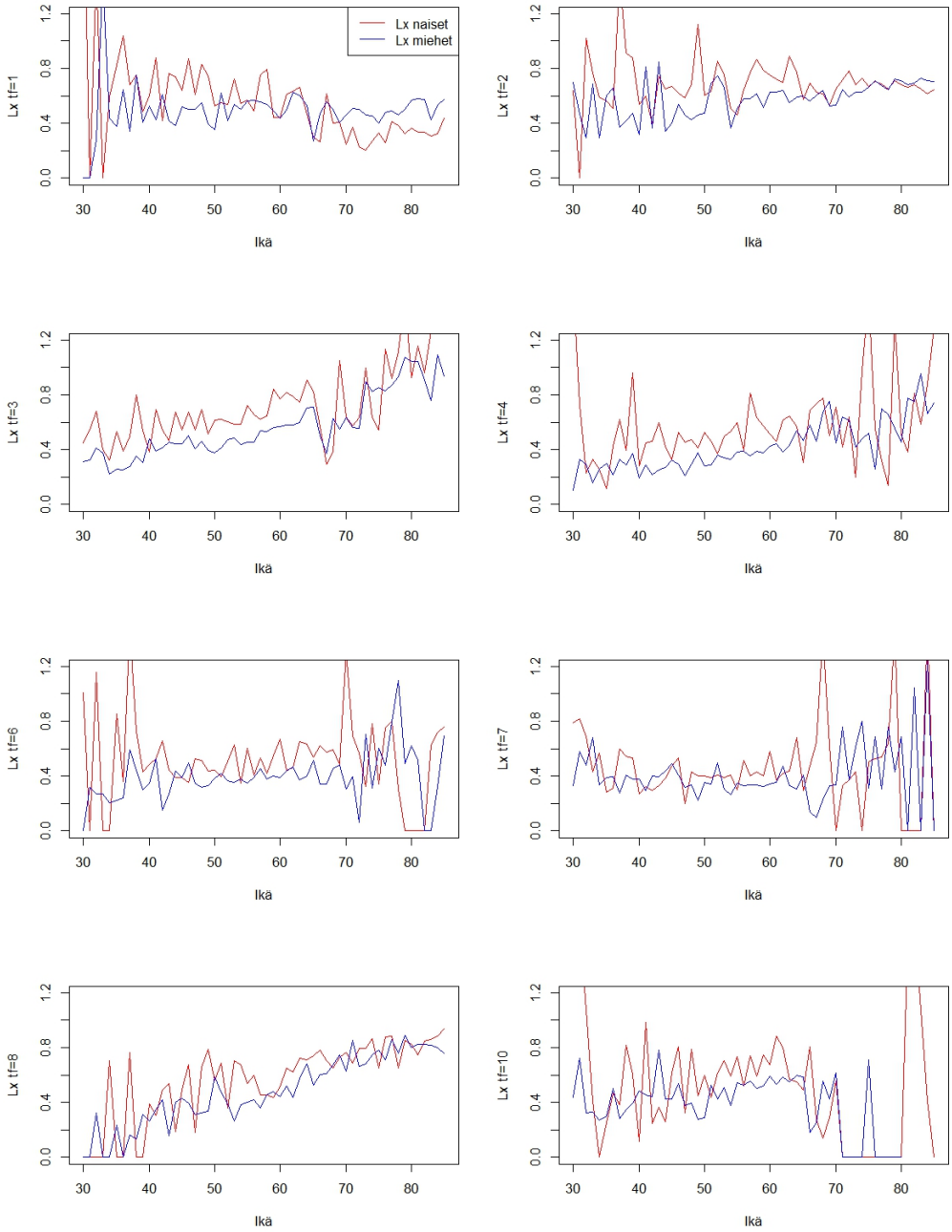
Seuraavissa tarkasteluissa esitetään lyhyesti seuraavat mallit:

Malli 1: $\ln(L_x) = \beta_0 + \beta_1 x$

Malli 2: $\ln(L_x) = \beta_0 + \beta_1 s$

Malli 3: $\ln(L_x) = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 sx$

Malli 4: $\ln(L_x) = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 sx + \beta_4 x^2$




Kuva 5.10: Havaitut Lx-luvut vakuutuslajeittain (tf1=yksilöllinen säästövakuutus, tf2=yksilöllinen säästövakuutus unit linked, tf3=yksilöllinen riskivakuutus, tf4=ryhmäetuvakuutus, tf6=yksilöllinen eläkevakuutus, tf7=yksilöllinen eläkevakuutus unit linked, tf8=ryhmäeläkevakuutus, tf10=tapaturmainen kuolema)

Malli 5: $\ln(L_x) = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 sx + \beta_4 x^2 + \beta_5 x^3$

Malleissa $x$ viittaa ikään, $s$ sukupuoleen (0= nainen, 1 = mies) ja kertoimet $\beta_j$ regressiokertoimiin. Malli 1 ei riipu sukupuolesta, joten se helpottaa naisten ja miesten erojen vertailua.

Poisson-regressiossa epävarmuus voidaan ottaa huomioon normaalisuusoletukseen nojautuen luottamusväleillä tai simulointimenetelmillä. Poisson-regressiossa parametrien estimointi on tehty suurimman uskottavuuden menetelmällä (SUE), jonka yhteydessä oletetaan yleensä multinormaalisuus. Teoriassa SUE:n jakauma lähestyy multinormaalijakaumaa, kun otoskoko lähestyy ääretöntä. Näin ollen suurimman uskottavuuden estimaattori vektorille $\boldsymbol{\beta}$ voidaan olettaa likimain normaalijakautuneeksi.

Luottamusvälit voidaan muodostaa seuraavasti. Oletetaan, että

$$\ln(\hat{L}_x) = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Tässä tapauksessa

$$\mathrm{Var}(\ln(\hat{L}_x)) = \mathrm{Var}(\hat{\beta}_0 + \hat{\beta}_1 x) = \mathrm{Var}(\hat{\beta}_0) + x^2 \mathrm{Var}(\hat{\beta}_1) + 2x\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1).$$

Kertoimien varianssit ja kovarianssi saadaan kovarianssimatriisista

$$\hat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}) = \left( \sum_{i=1}^{n} \hat{D}_i \mathbf{z_i} \mathbf{z_i}' \right)^{-1},$$

missä $\hat{\boldsymbol{\beta}}$ on parametrivektori, $\mathbf{z_i} = 1, x_{i1}$ on havaintoon $i$ liittyvä selittävien muuttujien vektori, n on havaintojen lukumäärä aineistossa ja $\hat{D}_i$ havaintoon $i$ liittyvä kuolleiden lukumäärän sovitettu arvo. Näin ollen normaalioletukseen perustuva $100(1-\alpha)\%$:n luottamusväli on

$$\ln(\hat{L}_x) \pm z_{\alpha/2} \hat{\sigma}_{\ln(L_x)}.$$

Toisin sanoen korjauskertoimen $\hat{L}_x$ luottamusväli on muotoa

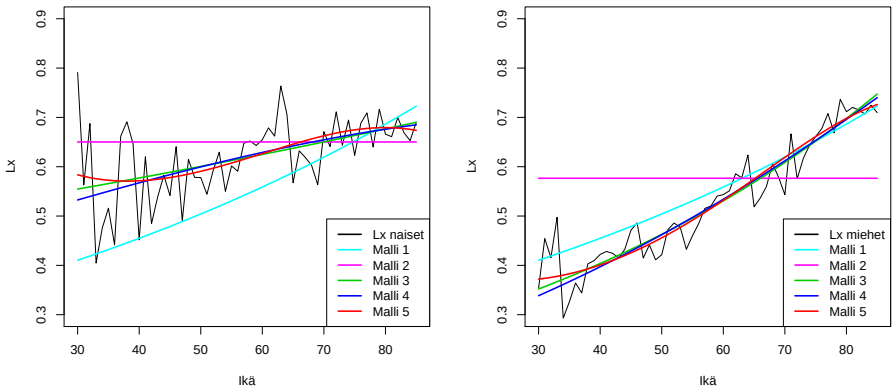
$$\hat{L}_x \exp(\pm z_{\alpha/2} \hat{\sigma}_{\ln(L_x)}).$$

### 5.6.2 Henkivakuutusaineisto

Tarkastellaan lyhyesti koko aineistoon liittyviä tuloksia. Näitä tuloksia voi käyttää myös vertailukohtana tuotekohtaisissa analyyseissä. Ikäväli on rajoitettu välille 30–85. Malleja on havainnollistettu kuvaajassa 5.11. Silmämääräisesti arvioiden toisen asteen termin lisääminen (malli 4) ei paranna mallia sovitettuun suoraan nähden (malli 3). Malli 5, jossa on otettu huomioon toisen ja kolmannen asteen termit, pystyy ottamaan hiukan paremmin huomioon nuoret ja vanhat iät.

Samoihin päätelmiin päädytään tutkimalla malleihin liittyviä tunnuslukuja. Mallissa 3 kaikki parametrit ovat merkitseviä, kun taas mallissa 4 toisen asteen termi ei ole merkitsevä. Mallissa 5 kolmannen asteen ja toisen asteen termit ovat merkitseviä. Näin ollen esitetyistä malleista vaikuttaisi olevan paras malli 3 tai malli 5. Mallin 3 Akaiken informaatiokriteeri (AIC) on 6036.8

kun taas mallin 5 AIC on 6033.6. Näin ollen Akaiken informaatiokriteerin perusteella malli 5 olisi hiukan parempi valinta (Brockwell & Davis (2002)). Kuitenkin tulee muistaa, että korjaustermiin sisältyy epävarmuutta sovitettuun malliin sekä taustalla olevan henkivakuutusaineistoon liittyen. Toisaalta myös korjaustermin oletetaan olevan sama koko ennustejakson ajan. Näin ollen tarkemman mallin valitseminen ei välttämättä ole merkitsevästi parempi käytännössä. Mallin valinnassa tuleekin käyttää asiantuntijaharkintaa. Toisaalta mallin epävarmuutta voidaan ottaa huomioon muun muassa luottamusvälien avulla.

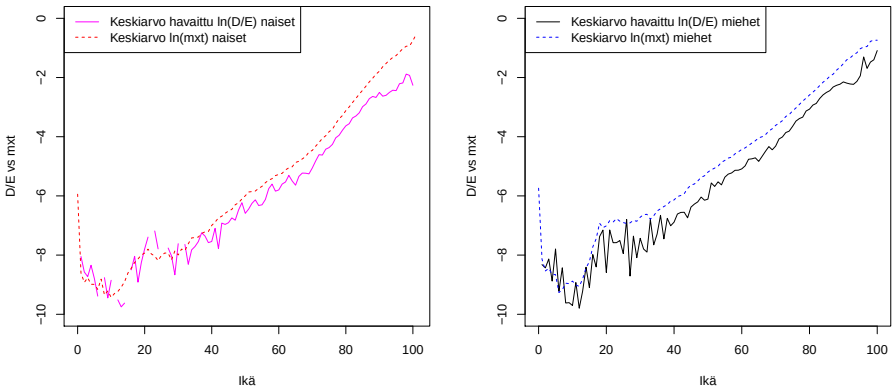



Kuva 5.11: Mallien sovittamat sekä havaitut korjauskertoimet $L_x$

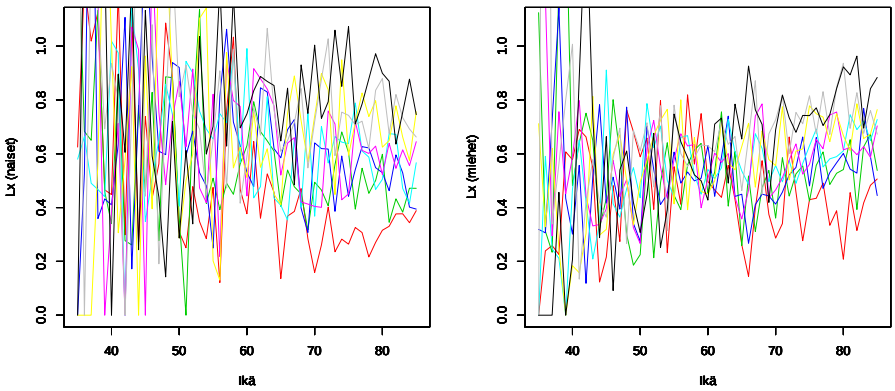
### 5.6.3 Yksilölliset säästövakuutukset

Säästöhenkivakuutuksissa vakuutettu säästää tulevaisuuden varalle. Näin ollen säästöhenkivakuutetut ovat usein keskimääräistä varakkaampia ja siten myös todennäköisesti keskimääräistä terveempiä. Toisaalta kuolevuus perustekorkoisissa- ja sijoitussidonnaisissa vakuutuksissa ei havaintojen mukaan poikkea toisistaan merkitsevästi (kuva 5.10). Vastuunvalinta näissä on samanlaista. Näin ollen näitä perustekorkoisia ja sijoitussidonnaisia säästövakuutuksia tutkitaan yhtenä ryhmänä seuraavissa tarkasteluissa. Tutkitaan siis seuraavaksi väestökuolevuuden ja säästöhenkivakuutettujen suhdetta.

Kuvassa 5.12 on havainnollistettu Suomen väestön keskimääräisiä kuolevuuksia $m_{x,t}$ ja säästöhenkivakuutusaineistosta havaittuja kuolevuuksia (kuolleiden lukumäärän suhde altistumaan eli "D/E-luku"). Silmämääräisesti arvioiden naisilla ero keskimääräiseen väestökuolevuuteen on pienempää kuin miehillä. Kuvasta 5.13 huomataan, että korjauskertoimen $L_{x,t}$ vaihtelu on todella suurta miehillä ja naisilla. Tähän vaikuttaa erityisesti vähäinen kuolleiden lukumäärä aineistossa. Lisäksi aineistossa on puutteita vuosina 2002 ja 2003. Seuraavissa tarkasteluissa ikäväliksi on valittu 35–85 ja aineiston puutteiden vuoksi vuosi 2002 on jätetty pois analyysista. Kun aineistoon on sovitettu luvussa 5.6.1 esitetyt mallit, mallien tuloksien perusteella voidaan päätellä, etteivät monimutkaisemmat mallit paranna mallia huomattavasti, joten valitaan malli 3. Malli on seuraavaa muotoa:




Kuva 5.12: Vuosilta 2002–2009 laskettu havaittu keskimääräinen kuolleiden lukumäärä jaettuna altistumalla sekä väestön keskimääräinen kuolevuus $m_x$ (yksilölliset säästövakuutukset, logaritmoitu)




Kuva 5.13: Miesten ja naisten $L_x$-suureet (yksilölliset säästövakuutukset)

$$\ln(L_x^{\text{miehet}}) = -1.10342 + 0.00870x$$

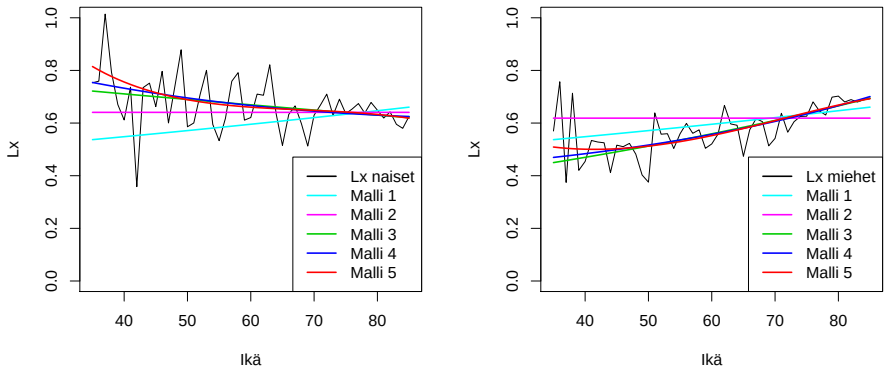
$$\ln(L_x^{\text{naiset}}) = -0.22098 - 0.00301x$$

Malliin 3 liittyvät luottamusvälit on esitetty kuvassa 5.15. Luottamusvälien kapeus selittyy sillä, että Poisson-regressiossa estimoidaan keskiarvokäyrä ja keskiarvon vaihtelu on paljon pienempi kuin alkuperäisten havaintojen vaihtelu.

### 5.6.4 Yksilöllinen riskivakuutus

Seuraavassa tarkastelussa tehdään analyysit erikseen yksilölliselle riskivakuutukselle sekä ryhmävakuutuksille, sillä vastuunvalinta on yleensä kevyempää ryhmävakuutuksissa. Näin ollen myös kuolevuus saattaa olla ryhmävakuutuksissa ja yksilöllisissä vakuutuksissa erilainen.

Yksilöllisessä riskivakuutuksessa kuolleiden suhde altistumaan vaikuttaa olevan erilainen naisilla ja miehillä. Lisäksi huomataan, että vanhemmissa ikä-




Kuva 5.14: Sovitetut mallit sekä havaittu keskimääräinen korjauskerroin (yksilölliset säästövakuutukset)

luokissa niin miehillä kuin naisilla väestökuolevuus $m_{x,t}$ ja havaittu keskikuolevuus ovat lähes samat (kuva 5.16). Tätä voidaan perustella muun muassa vastuunvalinnan vaikutuksen heikentymisellä vanhemmissa ikäluokissa.

Mallia sovitettaessa ikämuuttuja rajoitettiin välille 30–65. Miehillä suhdeluku väestökuolevuuteen on pienempi, eli yksilöllisesti riskivakuutettujen miesten kuolevuus eroaa miesten väestökuolevuudesta enemmän kuin naisten kuolevuus naisten väestökuolevuudesta (kuva 5.17). Lisäksi niin naisilla kuin miehillä korjauskerroin on ikäriippuvainen. Sovitetuista malleista paras oli suora eli malli 3. Miesten ja naisten korjauskertoimet määritellään seuraavasti:

$$\ln(L_x^{\text{miehet}}) = -1.95742 + 0.02318x$$

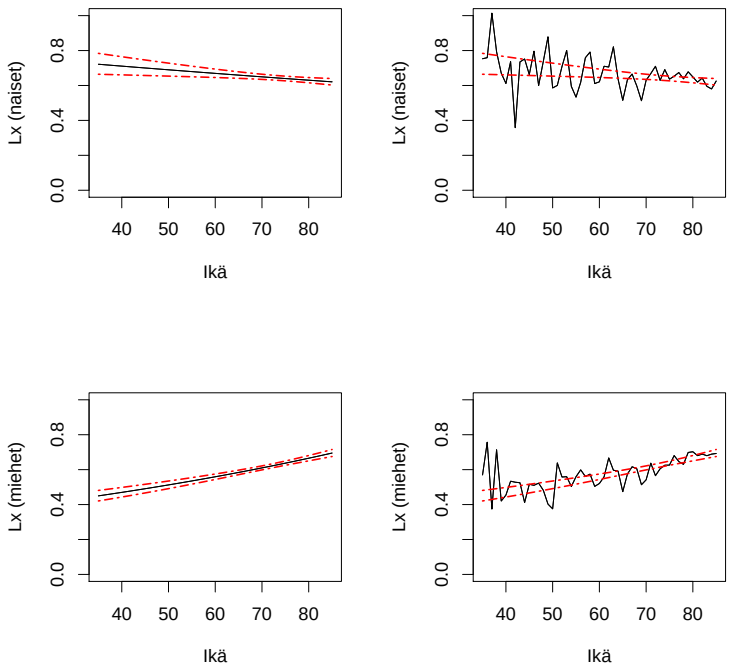
$$\ln(L_x^{\text{naiset}}) = -1.34338 + 0.01749x$$

Kuvassa 5.18 on esitetty K2004-referenssikuolevuuden ja yksilöllisten riskivakuutusten korjauskertoimet. Huomataan, että yksilöllisesti riskivakuutettujen korjauskerroin kulkee K2004-referenssikuolevuusmallin korjauskertoimien alareunan tuntumassa.

Liitteessä F on havainnollistettu Poisson-regression ja summakuolevuuden perusteella määriteltyjen korjauskertoimien eroja. Summakuolevuuden perusteella määritellyllä korjauskertoimella tarkoitetaan summakuolevuuden ja väestökuolevuuden suhteeseen $(\frac{m_x^S}{m_x^V})$ sovitettua regressiosuoraa.

### 5.6.5 Ryhmäetuvakuutus ja ryhmäriskivakuutus

Seuraavaksi tutkitaan ryhmäetu- sekä ryhmäriskivakuutuksia. Vastaavasti kuin yksilöllisessä riskivakuutuksessa vanhemmissa ikäluokissa havaittu keskimääräinen kuolevuus ja vastaava väestökuolevuus ovat lähellä toisiaan (kuva 5.19). Tässä tarkastelussa ikä on rajoitettu välille 35–65. Kuvassa 5.20 on kuvattu sovitetut mallit sekä havaittu keskimääräinen korjauskerroin. On mielenkiintoista huomata, että korjauskerroin on pienempi kuin yksilöllisessä




Kuva 5.15: Sovitettuun malliin liittyvät 95 % luottamusvälit (yksilölliset säästövakuutukset)

riskivakuutuksessa. Ryhmäetu- ja ryhmäriskivakuutusaineistojen suppeuden vuoksi tulokseen liittyy paljon epävarmuutta (taulukko 3.3). Kuitenkin tämä voi selittyä sillä, että asuntolainojen takaisinmaksun turvaksi otetaan usein vakuutus kuoleman varalta. Nämä vakuutukset sisältyvät yksilölliseen riskivakuutukseen. Ryhmäetu- ja ryhmäriskivakuutusten vakuuttajaryhmä poikkeaa niistä, jotka ottavat yksilöllisen riskivakuutuksen.

$$\ln(L_x^{\text{miehet}}) = -2.03520 + 0.02005x$$

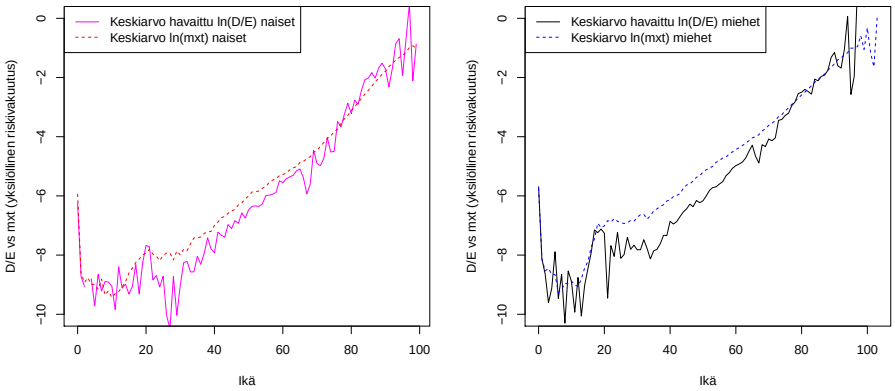
$$\ln(L_x^{\text{naiset}}) = -1.00706 + 0.00761x$$

Liitteessä F on nähtävissä Poisson-regression ja summakuolevuuden perusteella määriteltyjen korjauskertoimien ero.

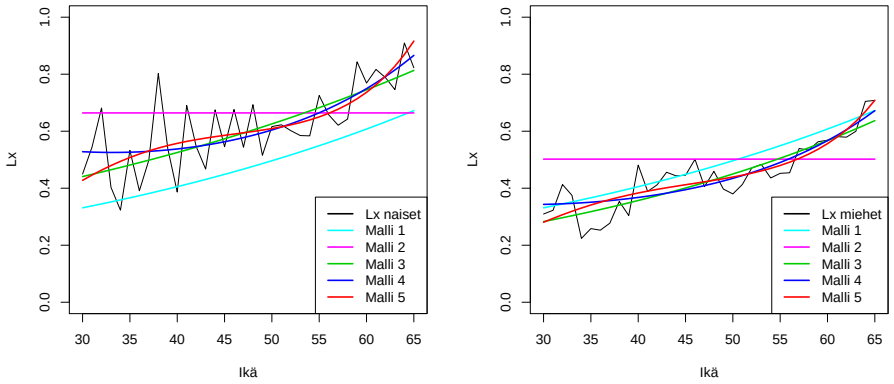
### 5.6.6 Yksilöllinen eläkevakuutus ja perhe-eläke

Yksilölliset eläkevakuutukset voivat olla perustekorkoisia vakuutuksia tai sijoitussidonnaisia (unit link) vakuutuksia tai näiden yhdistelmiä. Lisäksi tähän ryhmään on lisätty perhe-eläkkeet, minkä ansiosta aineistoon saadaan lisää erityisesti iäkkäitä naisia. Kuvasta 5.21 huomataan, että ero keskimääräiseen väestökuolevuuteen on suurempi miehillä kuin naisilla.

Kuvassa 5.22 on esitetty sovitetut mallit, joissa ikä on rajoitettu välille 35–70. Aineistoon sopii parhaiten suora. Alla on määritelty suorat miehille ja naisille.




Kuva 5.16: Vuosilta 2002–2009 laskettu havaittu keskimääräinen kuolleiden lukumäärä jaettuna altistumalla sekä väestön keskimääräinen kuolevuus $m_x$ (yksilöllinen riskivakuutus, logaritmoitu)




Kuva 5.17: Sovitetut mallit sekä havaittu keskimääräinen korjauskerroin yksilölliselle riskivakuutukselle

$$\ln(L_x^{\text{miehet}}) = -1.11171 + 0.00269x$$

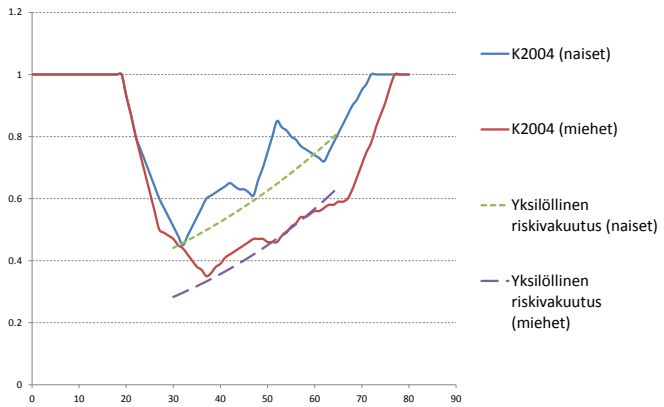
$$\ln(L_x^{\text{naiset}}) = -1.67621 + 0.01738x$$

### 5.6.7 Ryhmäeläkevakuutus

Ryhmäeläkkeissä niin naisilla kuin miehilläkin havaittu kuolevuus suhteessa väestökuolevuuteen on lähes sama vanhemmissa ikäluokissa (kuva 5.23). Mallia sovitettaessa ikä on rajoitettu välille 40–80. Tilastollisten testien perusteella suora sopii aineistoon parhaiten (kuva 5.24).

$$\ln(L_x^{\text{miehet}}) = -2.37356 + 0.02791x$$

$$\ln(L_x^{\text{naiset}}) = -1.43490 + 0.01577x$$




Kuva 5.18: K2004-referenssikuolevuuden korjauskertoimet (Mäkinen (2004)) ja yksilöllisen riskivakuutuksen korjauskertoimet

Luvussa 5.2 havaittiin, että ryhmäeläkevakuutuksissa miesten summakuolevuus poikkesi lukumääräkuolevuudesta (kuva 5.1). Tästä syystä myös rahastokuolevuuden ja väestökuolevuuden suhteen ($\frac{m_x^S}{m_x^V}$) perusteella määriteltiin korjauskerroin $L_x^S$ regressioanalyysillä. Toiseen asteen polynomi sopi aineistoon parhaiten. Miesten summakuolevuuteen perustuva korjauskerroin on muotoa:

$$L_x^{\mathrm{S,\ miehet}} = 1.18087 - 0.03980x + 0.00043x^2$$

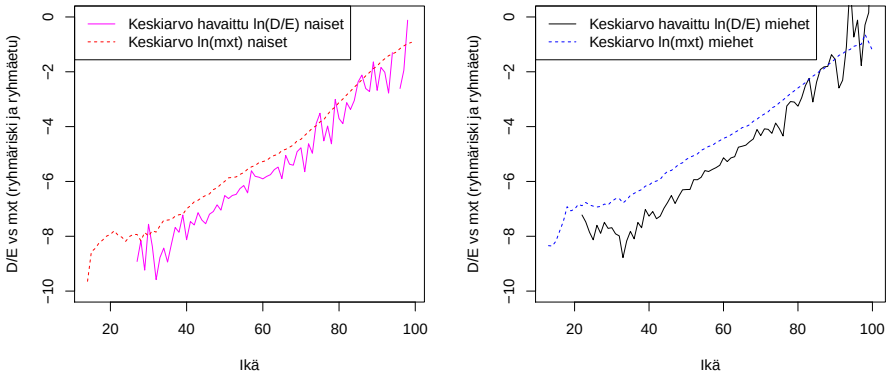
Kuvassa 5.25 on esitetty summakuolevuuteen perustuvat sovitteet sekä edellä esitetty Poisson-regressioon perustuva korjauskerroin. Huomataan, että lukumääräaineistoon perustuva korjauskerroin ei pysy summakuolevuuteen perustuvan korjauskertoimen 95 % luottamusvälin sisäpuolella. Ero on suurinta 60 ikävuoden tienoilla.

### 5.6.8 Tapaturmainen kuolema

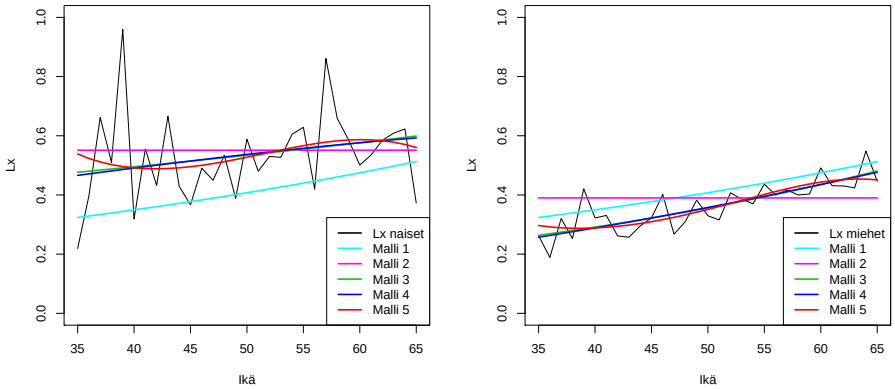
Vertailtaessa D/E-lukua väestön keskikuolevuuksiin huomataan, että erityisesti miehillä havaittu kuolevuus on kaikissa ikäluokissa alempana kuin keskimääräinen väestökuolevuus (kuva 5.26). Naisilla ero on hiukan pienempi. Malleista suora tai vakiokerroin sopii parhaiten aineistoon (kuva 5.27). Mallissa ikäväli on rajoitettu välille 35–65.

$$\ln(L_x^{\mathrm{miehet}}) = -1.67407 + 0.01776x$$

$$\ln(L_x^{\mathrm{naiset}}) = -1.18685 + 0.01301x$$



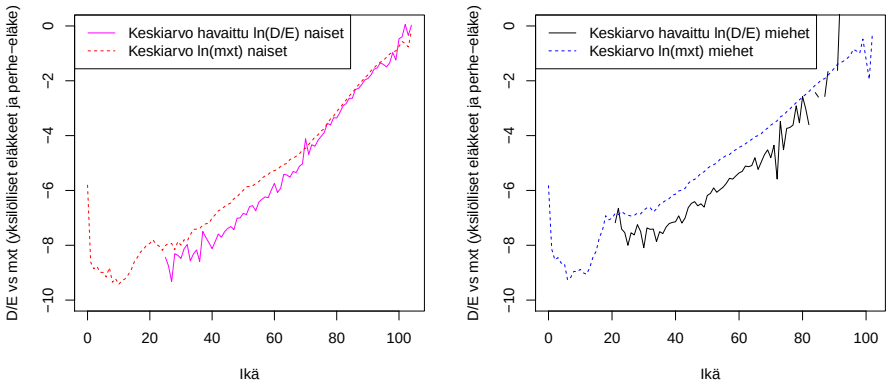

Kuva 5.19: Vuosilta 2002–2009 laskettu havaittu keskimääräinen kuolleiden lukumäärä jaettuna altistumalla sekä väestön keskimääräinen kuolevuus $m_x$ (ryhmäriskivakuutus ja ryhmäetuvakuutus, logaritmoitu)



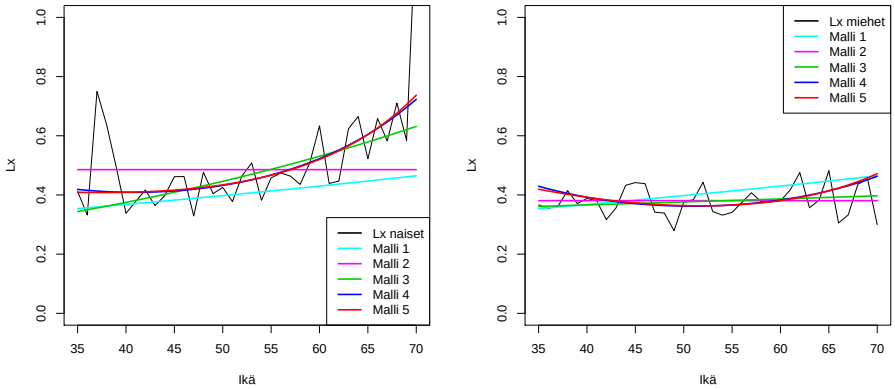

Kuva 5.20: Sovitetut mallit sekä havaittu keskimääräinen korjauskerroin (ryhmäetuvakuutus ja ryhmäriskivakuutus)

## 5.7 Yhteenveto

Summa- ja lukumääräkuolevuus erosivat toisistaan erityisesti miehillä, kun taas naisilla kuolevuudet olivat lähellä toisiaan. Miehillä eroja syntyi eniten ryhmäeläkevakuutuksessa. Myös riskivakuutuksissa ja yksilöllisessä eläkevakuutuksessa oli eroja miesten osalta. Kuitenkin luvussa 5.2 summa- ja lukumääräkuolevuuden havaittiin olevan lähes kaikissa tapauksissa riittävän lähellä toisiaan ottaen huomioon tämän raportin tarkoitus määrittää keskimääräinen henkivakuutuskuolevuus tuotteittain ja korjauskertoimiin ja kuolevuusennusteeseen liittyvä epävarmuus. Näin ollen tässä raportissa korjauskertoimet kaikille tuotteille määriteltiin lukumääräaineiston pohjalta Poisson-regressiolla. Lisäperustelu lukumääräaineistoon keskittymiseen oli muun muassa volyymiaineiston puutteellisuus. Kuitenkin muun muassa miesten ryhmäeläkkeille esitettiin myös summakuolevuuteen perustuvat korjauskertoi-




Kuva 5.21: Vuosilta 2002–2009 laskettu havaittu keskimääräinen kuolleiden lukumäärä jaettuna altistumalla sekä väestön keskimääräinen kuolevuus $m_x$ (yksilölliset eläkevakuutukset ja perhe-eläke, logaritmoitu)

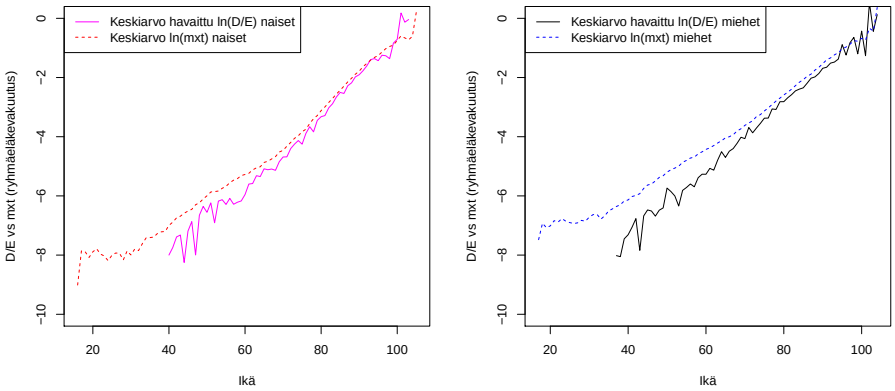



Kuva 5.22: Sovitetut mallit sekä havaittu keskimääräinen korjauskerroin (yksilölliset eläkevakuutukset ja perhe-eläke)

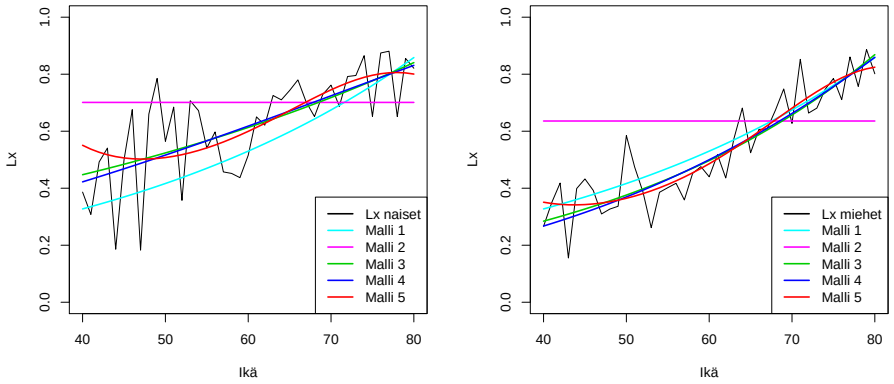
met, koska summa- ja lukumääräkuolevuudet poikkesivat toisistaan selvästi.

Kuvissa 5.28–5.29 on esitetty lukumääräaineistosta Poisson-regressiolla määritellyt korjauskertoimet vakuutuslajeittain. Tuotteittaiset korjauskertoimet on voitu määritellä pääosin ikävälille 35–65. Miehillä vakuutuslajikohtaiset korjauskertoimet ovat pienempiä kuin naisilla. Naisilla korjauskerroin on pienimmillään yksilöllisessä eläkevakuutuksessa. Vertailtaessa ryhmäeläkkeitä ja yksilöllisiä eläkevakuutuksia sukupuolittain huomataan, että ryhmäeläkkeiden kuolevuus on lähempänä väestökuolevuutta erityisesti vanhemmissa ikäluokissa.

Luvussa 5.6.4 todettiin, että yksilöllisen riskivakuutuksen korjauskertoimet kulkevat aikaisemman tutkimuksen K2004-korjauskertoimien alarajalla (kuvio 5.18). Vertailtaessa TyEL:n rahastokuolevuuteen perustuvia korjauskertoimia ja tämän tutkimuksen eläkevakuutusten lukumääräkuolevuuteen perustuvia korjauskertoimia huomataan, että TyEL:n korjauskerroin on sekä




Kuva 5.23: Vuosilta 2002–2009 laskettu havaittu keskimääräinen kuolleiden lukumäärä jaettuna altistumalla sekä väestön keskimääräinen kuolevuus $m_x$ (ryhmäeläkevakuutus, logaritmoitu)

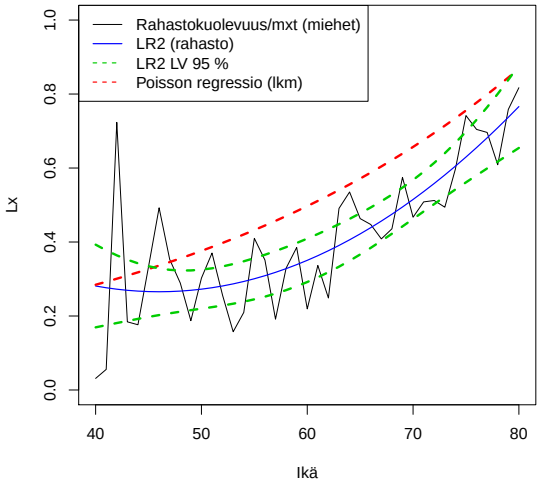



Kuva 5.24: Sovitetut mallit sekä havaittu keskimääräinen korjauskerroin (ryhmäeläkevakuutus)

naisilla että miehillä korkeampi (kuvio 5.30, Hakkarainen (2012)). Tämä on odotettua henkivakuutuksiin liittyvän selektin vuoksi.

Esitetyt korjauskertoimet kuvaavat henkivakuutuskannan keskimääräistä henkivakuutuskuolevuuden ja väestökuolevuuden suhdetta. Niin kuin aikaisemmin on todettu, yhtiöiden välillä vastuunvalinta voi poiketa ja muun muassa asiakaskanta ja sopimukset ovat erilaisia. Toisaalta myös aineiston koko ja laatu vaikuttavat korjauskertoimiin liittyvään epävarmuuteen. Näin ollen jos yhtiö laskee omat korjauskertoimensa, silloin tulisi ottaa huomioon yhtiön vakuutuskannan ominaisuudet ja aineiston rajoitteet. Laskelman voi tehdä esimerkiksi sovittamalla yhtiökohtaiseen aineistoon Poisson-regression (R-ohjelman ohje liitteessä H) tai muulla tavoin arvioiden.

Lukumääräaineistoon perustuvia korjauskertoimia varten havaintovuosilta lasketaan kuolleiden lukumäärä $D_{x,t}$ ja odotettujen kuolleiden lukumäärä $E_{x,t}m_{x,t}$ (vertaa luku 5.5). Tämän jälkeen korjauskerroin voidaan laskea kaa-




Kuva 5.25: Rahastokuolevuuden ja väestökuolevuuden suhteeseen sovitettu toisen asteen polynomi (LR2) regressioanalyysillä. Lisäksi 95 %:n luottamusväli LR2:lle sekä Poisson regressioon perustuva korjauskerroin. (Ryhmäeläke, miehet)

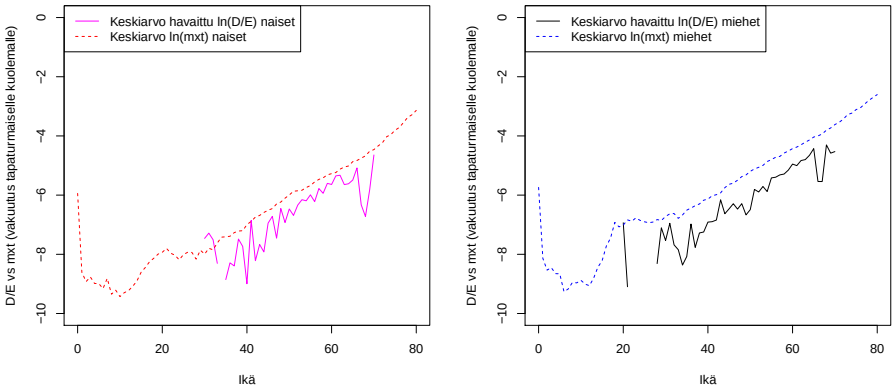
valla

$$L_x = \frac{\sum_t D_{x,t}}{\sum_t m_{x,t} E_{x,t}}.$$

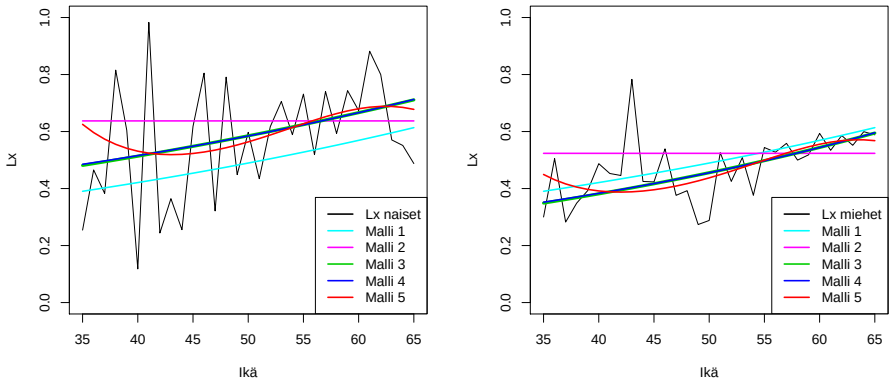
Korjauskerroin oletetaan riippumattomaksi ajasta, joten korjauskertoimen oletetaan pysyvän vakiona koko ennusteperiodin ajan.

Laskelmia tehtäessä tulee ottaa huomioon aineiston riittävyys ja järkevyys. Aineiston järkevyydellä tarkoitetaan, että vakuutuskanta ja vastuunvalinta eivät ole muuttuneet merkitsevästi dataperiodin aikana. Aineiston riittävyyttä voi arvioida muun muassa kuvaajien (esimerkiksi 5.9 tai 5.27) sekä odotettujen kuolleiden ja altistuneiden lukumäärien avulla. Mitä pienempi tarkasteltava kanta on, sitä pidempi aikaperiodi tulisi valita. Jos aineistoa on tarpeeksi, korjauskertoimet voi laskea myös eri aikaperiodeilta ja verrata tuloksia toisiinsa. Tällä tavoin voi testata oletusta, että korjauskertoimet eivät riipu ajasta.

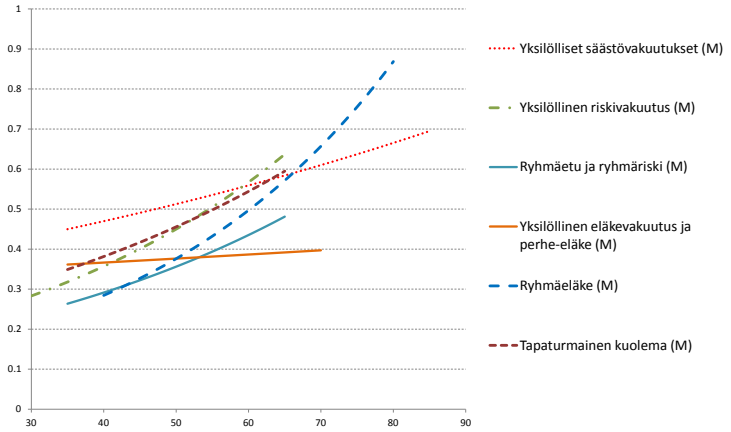
Kuolevuusperustetta tehtäessä tulee ottaa myös kantaa korjauskertoimen suuruuteen ikävälille, jolle korjauskerrointa ei voida estimoida havaintoaineiston perusteella. Ekstrapoloinnissa tulisi ottaa huomioon tuotteiden ja vakuutuskannan ominaisuudet, joten ekstrapolointi vaatii siis asiantuntijaharkintaa.



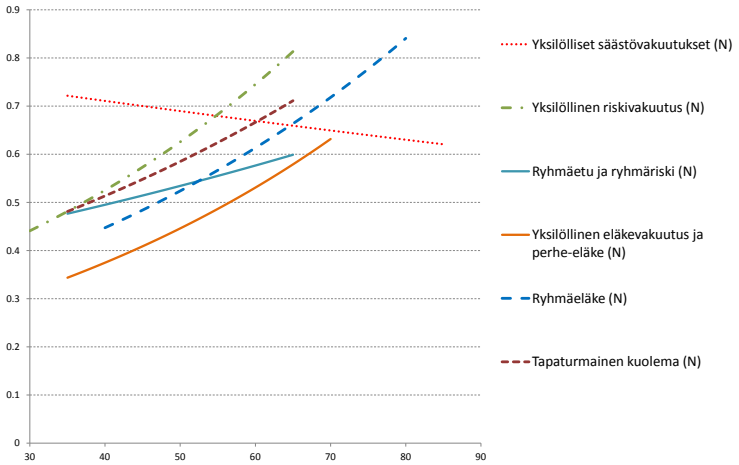

Kuva 5.26: Vuosilta 2002–2009 laskettu havaittu keskimääräinen kuolleiden lukumäärä jaettuna altistumalla sekä väestön keskimääräinen kuolevuus $m_x$ (tapaturmainen kuolema, logaritmoitu)



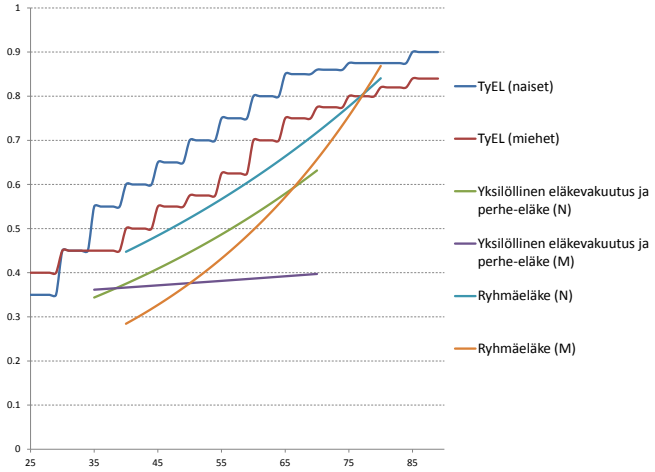

Kuva 5.27: Sovitetut mallit sekä havaittu keskimääräinen korjauskerroin (tapaturmavakuutus)




Kuva 5.28: Miesten korjauskertoimet vakuutuslajeittain




Kuva 5.29: Naisten korjauskertoimet vakuutuslajeittain




Kuva 5.30: Eläkevakuutusten korjauskerroin verrattuna TyEL-tutkimuksen kertoimeen

# 6 Raportin yhteenveto

Suomen väestön 120 vuoden kuolevuusennuste tehtiin Lee-Miller -menetelmällä käyttäen Suomen väestöaineistona HMD-tietokannan tasoitettuja tilastoja vuosilta 1955–2009. Ennusteessa ei otettu huomioon kuolevuuden shokkityyppistä kasvua. Ennuste tehtiin 120 vuodelle, jotta aineistosta on mahdollista muodostaa kohorttiaineisto. Kuolevuusennuste tehtiin sukupuolittain ikävälille 0–110+. Ennusteen epävarmuus otettiin huomioon normaalijakaumaoletukseen perustuvilla luottamusväleillä. Erona K2004-referenssikuolevuuteen tässä raportissa on Lee-Carter -menetelmän Lee-Miller -muunnelma sekä lähtöaineisto, joka muuttui vuosista 1960–2001 vuosiin 1955–2009. Lisäksi tässä raportissa on otettu huomioon ennusteen epävarmuus.

Henkivakuutuskuolevuus määriteltiin korjauskertoimella Suomen väestön kuolevuusennusteesta. K2004-referenssikuolevuudessa korjauskerroin määriteltiin yksilöllisen henkivakuutuksen summakuolevuuden perusteella (aineisto vuosilta 1972–2001). Tässä raportissa korjauskerroin määriteltiin vakuutuslajiryhmittäin Poisson-regressiolla lukumääräkuolevuuksien perusteella (henkivakuutusaineisto vuosilta 2002–2009). Kuitenkin miesten ryhmäeläkevakuutuksille tuotettiin myös summakuolevuuteen perustuva korjauskerroin, koska summa- ja lukumääräkuolevuudet poikkesivat toisistaan. Lukumääräaineiston painottamiseen päädyttiin muun muassa aineiston laatuun liittyvistä syistä. Lisäksi sen todettiin olevan tässä tarkoituksessa riittävän lähellä summakuolevuutta lähes kaikissa vakuutuslajeissa (luku 5.2), kun otetaan huomioon korjauskertoimiin ja kuolevuusennusteeseen liittyvä epävarmuus. K2004-referenssikuolevuudessa ja tässä raportissa oletettiin, että korjauskerroin pysyy vakiona koko ennusteperiodin ajan.

Henkivakuutusaineisto sisältää kymmenen eri vakuutuslajia. Vakuutuslajit eroavat toisistaan muun muassa vastuunvalinnaltaan ja vakuutusriskiltään. Toisin sanoen henkivakuutusaineisto on heterogeeninen ja vakuutuslajikohtaiset tarkastelut ovat perusteltuja. Kuitenkin havaintojen ja kuolintapauksien vähäisyyden vuoksi estimoituihin korjauskertoimiin sisältyy epävarmuutta. Korjauskertoimet vaihtelevat myös yhtiöittäin.

Raportissa verrattiin saatuja tuloksia aiempiin vastaaviin tutkimuksiin. Luvussa 5.6.4 havaittiin, että K2004-tutkimuksen korjauskertoimet ovat hiukan korkeammalla tasolla kuin yksilöllisen riskivakuutuksen korjauskertoimet (kuva 5.18). TyEL:n summakuolevuuteen perustuva korjauskerroin on puolestaan korkeammalla tasolla kuin eläkevakuutuksien korjauskertoimet (kuva 5.30, Hakkarainen (2012)).

Liitteessä D on esitetty väestöennusteen kuolevuudet ja niihin liittyvät hajonnat. Halutun turvaavuustason luottamusvälin saa laskettua luvussa 4.4 esitetyllä kaavalla. Luvussa 5.6 on esitetty vakuutustuotteiden korjauskertoimet. Korjauskertoimet on esitetty pääosin ikävälille 35–65. Kuolevuuspe-

Taulukko 6.1: Korjauskertoimen vaikutus elinajanodotteisiin (yksilöllinen eläkevakuutus ja perhe-eläke)

| Yksilöllinen eläke ja perhe | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Naiset | | | | Miehet | | | |
| Syntymävuosi | 1970 | 1960 | 1950 | 1940 | 1970 | 1960 | 1950 | 1940 |
| Ikä | 40 | 50 | 60 | 70 | 40 | 50 | 60 | 70 |
| Elinajanodote (väestö) | 49.34 | 38.54 | 28.18 | 18.44 | 42.80 | 32.38 | 22.85 | 14.62 |
| Elinajanodote (korjaus) | 51.10 | 40.21 | 29.57 | 19.37 | 47.28 | 36.68 | 26.48 | 16.91 |
| Erotus | 1.76 | 1.68 | 1.39 | 0.93 | 4.48 | 4.30 | 3.64 | 2.29 |
| % | 3.6 % | 4.4 % | 4.9 % | 5.0 % | 10.5 % | 13.3 % | 15.9 % | 15.7 % |

Taulukko 6.2: Korjauskertoimen vaikutus elinajanodotteisiin (ryhmäeläke)

| Ryhmäeläke | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Naiset | | | | Miehet | | | |
| Syntymävuosi | 1970 | 1960 | 1950 | 1940 | 1970 | 1960 | 1950 | 1940 |
| Ikä | 40 | 50 | 60 | 70 | 40 | 50 | 60 | 70 |
| Elinajanodote (väestö) | 49.34 | 38.54 | 28.18 | 18.44 | 42.80 | 32.38 | 22.85 | 14.62 |
| Elinajanodote (korjaus) | 50.81 | 39.91 | 29.30 | 19.20 | 45.93 | 35.14 | 24.86 | 15.65 |
| Erotus | 1.41 | 1.33 | 1.08 | 0.72 | 3.11 | 2.74 | 2.00 | 1.02 |
| % | 2.9 % | 3.5 % | 3.8 % | 3.9 % | 7.3 % | 8.5 % | 8.8 % | 7.0 % |

rusteessa tulee ottaa kantaa myös vanhempien ikien korjauskertoimiin (luku 5.7). Henkivakuutuskuolevuuden tietylle vakuutustuotteelle saa väestökuolevuuden $m_{x,t}$ ja tuotteen korjauskertoimen $L_x$ tulona. Liitteessä E on esitetty, miten henkivakuutuskuolevuuden $m^H_{x,t}$ epävarmuuden voi ottaa huomioon luottamusvälien avulla.

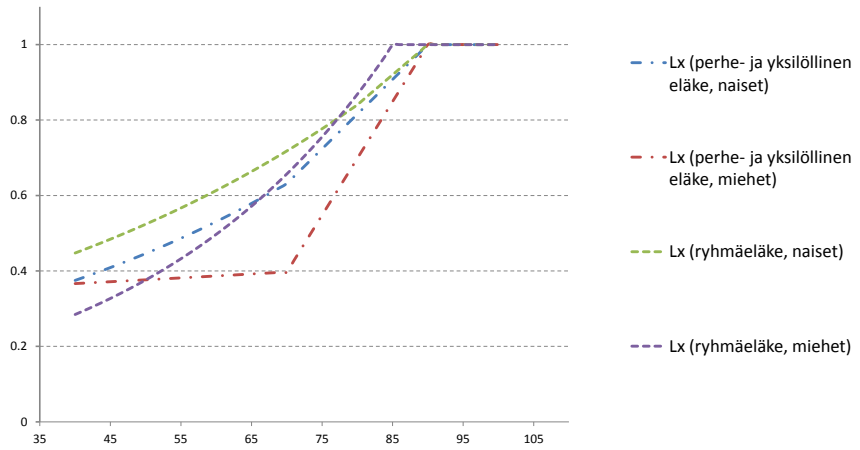
Liitteessä H on tutkimuksessa käytetyn R-ohjelmiston käyttöohje. Ohjeessa on listattu keskeisimmät komennot, joita tutkimuksessa käytettiin. Ohjeiden tavoitteena on mahdollistaa tutkimuksen toistettavuus.

Taulukoissa 6.1 ja 6.2 on esitetty korjauskertoimen vaikutus elinajanodotteisiin eläkevakuutuksissa. Elinajanodotteet on laskettu kohorteittain (liite D). Kuvassa 6.1 on esitetty laskelmassa käytetyt korjauskertoimet (vrt. lukumääräaineistoon perustuvat korjauskertoimet). Miehillä korjauskertoimen vaikutus on odotetusti suurempi.

Kuvissa 6.2 ja 6.3 on esitetty sukupuolittain yksilöllisen riskivakuutuksen kuolevuusennusteen sekä K2004-sovitteen suhteellinen ero kohorteille 2010–2019 ($(m^{HK2012}_{x,t} - m^{K2004}_{x,t})/m^{K2004}_{x,t}$). Korjauskertoimina on käytetty luvussa 5.6.4 esitettyjä korjauskertoimia. Ikäväleille, joille korjauskertoimia ei ole määritelty, on käytetty K2004-korjauskertoimia. Näin ollen ikäväleillä 0–29 ja 66–110 suhteellinen ero selittyy täysin väestöennusteen muutoksella ja sovitteen arviovirheellä. Myös ikävälillä 30–65 suuri osa erosta selittyy näillä tekijöillä, koska K2004-korjauskertoimet ja yksilöllisen riskivakuutuksen korjauskertoimet ovat lähellä toisiaan (kuva 5.18).

Naisilla suurin ero on ikävälillä 30–40 sekä 65–85 (kuva 6.2). Miehillä ero on suurinta ikäväleillä 25–45, 65–85 sekä 50 ikävuoden tienoilla (kuva 6.3). Naisilla ja miehillä suhteellisen erotuksen taso vaihtelee eri ikäluokissa. Verrattaessa kuvaa 6.2 kuvaan 6.3 voidaan todeta, että K2004-sovite sopii miehille paremmin kuin naisilla.

Liitteessä G on esitetty vastaava tarkastelu K2004-tutkimukselle sukupuo-



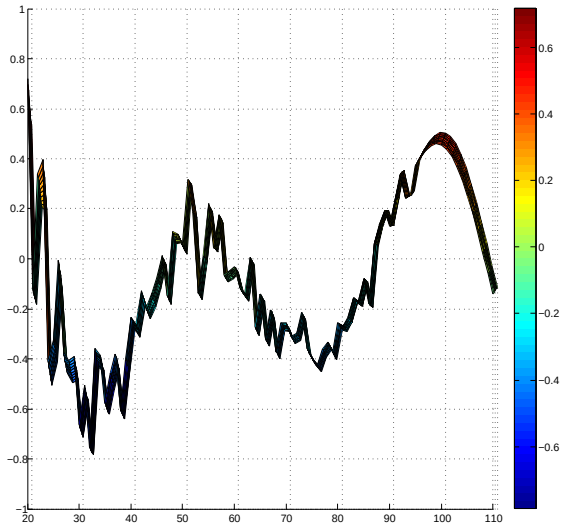

Kuva 6.1: Arviolaskelmassa käytetyt korjauskertoimet

littain. Tarkastelussa on vertailtu K2004-sovitetta sekä K2004-korjauskertoimilla korjattua Lee-Carter -ennustetta (lähtövuodet 1960–2001) toisiinsa. Sovitetta on vaikea saada ennustetta vastaavaksi koko ikävälillä (Liite G). Näin ollen K2004-sovite on pyritty saamaan mahdollisimman tarkaksi ikävälille 30–60 (luku 2.2). Naisilla arviovirhe on suurempaa kuin miehillä.

Tulosten perusteella voidaan harkita, tulisiko K2004-sovitteen parametrit määritellä erikseen eri tuotteille. Kolehmainen (2009) on tutkinut SHV-työssään asiaa vapaamuotoisen työeläkevakuutuksen tapauksessa. Tutkimuksessa päädyttiin silloisen aineiston perusteella korjaamaan K2004-funktiota kasvattamalla miesten syntymävuotta kuudella vuodella ja alentamalla ikää kolmella vuodella. Naisilla syntymävuotta korotettiin kahdella vuodella ja alentava ikäsiirto oli kaksi vuotta. Voitaisiin myös etsiä sellaista sovitefunktiota, joka annetulla kriteerillä sopii aineistoon mahdollisimman hyvin. Voidaan myös harkita, että käytetään suoraan korjauskertoimella korjattuja liitteen D $m_{x,t}$-lukuja halutulla luottamustasolla tuotteittain.

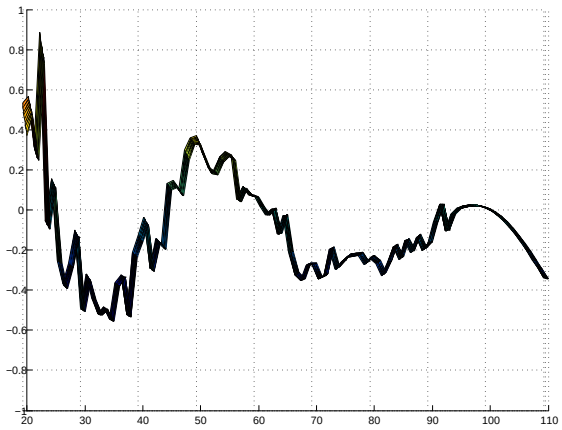
Koska kuolevuustutkimuksia tullaan toteuttamaan jatkossakin, yhtiöiden kannattaa varautua aineiston toimittamiseen. Tutkimuksessa tarvitaan kuolleiden lukumäärä ja altistuneiden lukumäärä sekä vastaavat volyymisuureet summatasolla vuosittain, sukupuolittain, tuotteittain ja ikäluokittain. Näin voidaan laskea korjauskertoimet luvussa 5 esitetyllä tavalla tuotteittain.

Mahdollisia jatkotutkimuksen aiheita ovat muun muassa analyyttisen sovitteen määrittäminen jollekin tuotteelle sekä simulointialgoritmi, jossa otetaan huomioon jakaumaoletukset sekä korjauskertoimiin liittyvä epävarmuus. Näin voitaisiin myös vertailla simulointialgoritmilla ja eri luottamusväleillä saatuja tuloksia toisiinsa. Lee-Miller mallin rajoitteita voidaan poistaa käyttämällä yleisempiä kuolevuusmalleja kuten esimerkiksi epälineaarisia malleja




Kuva 6.2: K2004-sovitteen ja yksilöllisen riskivakuutuksen henkivakuutuskuolevuuden suhteellinen ero kohorteille 2010-2019 (naiset).

(Alho et al. (2012)) tai lokaaleja malleja (Ronkainen (2012)). Lisäksi mallia voisi laajentaa sisältämään kuolevuuden shokit.




Kuva 6.3: K2004-sovitteen ja yksilöllisen riskivakuutuksen henkivakuutuskuolevuuden suhteellinen ero kohorteille 2010-2019 (miehet).

# Lähteet


Alho, M., J. (1990), ”Stochastic methods in population forecasting”, *International Journal of Forecasting*, 6, 521–530.

Alho, M., J., Bravo, J. & Palmer, E. (2012), ”Annuities and Life Expectancy in NDC”, teoksessa *Pension Reform II*, The World Bank.

Alho, M., J. & Spencer, D., B. (2005), *Statistical demography and forecasting*, New York: Springer.

Booth, H., Hyndman, R. & Piet, J. (2006), ”Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions”, *Demographic Research Volume 15, Article 9, pages 289–310.*

Booth, H. & Tickle, L. (2008), ”Mortality modelling and forecasting: A review of methods”, *ADSRI Working Paper No.3.*

Brockwell, P. J. & Davis, R. A. (2002), *Introduction to time series and forecasting* (2nd ed.), New York: Springer.

Cairns, A., Blake, D. & Dowd, K. (2008), ”Modelling and management of mortality risk: a review”, *Scandinavian Actuarial Journal*, pp. 79–113.

Crawley, M. (2007), *The R Book*, England: Wiley.

Dobson, A. (2002), *An Introduction to Generalized Linear Models*, USA: Chapman & Hall/CRC.

Flourens, P. (1850), *Histoire des travaux et des ides de Buffon* (2nd ed.), Paris.

Gaille, S. & Sherris, M. (2010), ”Modeling Long-Run Cause of Death Mortality Trends”, *Australian School of Business Research Paper No. 2010ACTL14.* Available from Internet: `http://www.asb.unsw.edu.au/schools/actuarialstudies/Documents/S.%20Gaille%20and%20M.%20Sherris%20-%20Modeling%20Long-Run%20Cause%20of%20Death%20Mortality%20Trends.pdf`.

Gerber, H. (1990), *Life Insurance Mathematics*, Berlin Heidelberg: Springer-Verlag.

Girosi, F. & King, G. (2008), *Demographic forecasting*, Princeton University Press.

Greider, C. & Blackburn, E. (1985), ”Identification of a Specific Telomere Terminal Transferase Activity in Tetrahymena Extracts”, *Cell*, 43, 405–413.

Hakkarainen, T. (2012), ”TyEL-kuolevuuden ennustamisesta”, SHV-työ.

Holzmann, R. & Palmer, E. e. (2006), *Pension Reform*, Washington, USA: The World Bank.

IAA (2010), *Stochastic Modeling, Theory and reality from an actuarial perspective*, Ottawa, Canada: International Actuarial Association.

Kaas, R., Goovaerts, J., M. Dhaene & Denuit, M. (2008), *Modern Actuarial Risk Theory*, Berlin: Kluwer Academic Publishers.

Kauppi, M. (2004), ”Lakisääteisen tapaturmavakuutuksen referenssikuolevuus”, Tapaturmavakuutuslaitosten liitto.

Keilman, N. (1990), *Uncertainty in national population forecastsing: Issues, backgrounds, analyses, recommendations.*, Amsterdam: Swets and Zeitlinger.

Keilman, N., Cruijsen, H. & Alho, M., J. (2008), ”Changing views of future demographic trends”, teoksessa *Uncertain demographics and fiscal sustainability*, Cambridge University Press.

Kolehmainen, P. (2009), ”Vapaamuotoisen työeläkevakuutuksen kuolevuus”, SHV-työ.

Kuusela, M. & Kukkala, M. (2010), ”Ennuste Suomen kuolevuuden kehityksestä”, Silta Oy.

Lee, R. & Carter, L. (1992), ”Modeling and Forecasting U.S. Mortality”, *Journal of the American Statistical Association*, 87, 659–671.

Lee, R. & Miller, T. (2001), ”Evaluating the performance of the Lee-Carter method for forecasting mortality.”, *Demography*, 38, 537–549.

Mäkinen, M. (2004), ”Referenssikuolevuus henkivakuutusyhtiöille”, SHV-työ.

Oeppen, J. & Vaupel, J. (2002), ”Broken Limits to Life Expectancy”, *Science*, 296, 1029–1031.

ONS (2009), ”Health Quarterly, Summer”, Office for national statistics.

PricewaterhouseCoopers (2011), ”Lakisääteisen tapaturmavakuutuksen ja liikennevakuutuksen kuolevuusreferenssimalli”, PricewaterhouseCoopers.

Rau, R., Soroko, E., Jasilionis, D. & Vaupel, J. (2006), ”10 years after Kannisto: further evidence for mortality decline at advanced ages in developed countries”, MPIDR Working Paper. Available from Internet: `http://www.demogr.mpg.de/papers/working/wp-2006-033.pdf`.

——— (2008), ”Continued Reductions in Mortality at Advanced Ages”, 34.

Robine, J., Roy, L., Jagger, C. & EHEMU (2005), ”Changes in life expectancy in the European Union since 1995: similarities and differences between the 25 EU countries”, EHEMU report (November). Available from Internet: `http://www.eurohex.eu/pdf/JM_Budapest.pdf`.

Ronkainen, V. (2012), *Stochastic modeling of financing longevity risk in pension insurance*, Helsinki: Suomen Pankin julkaisut.

Sandström, A. (2006), *Solvency: Models, Assessment and Regulation*, Chapman and Hall/CRC.

SHY (1982), *Henki- ja eläkevakuutuksen vakuutustekniikka.*

# Liite A

# Aineiston kuvaus

Henkivakuutusyhtiöiden kuolevuusselvitystä varten Finanssivalvonta keräsi henkivakuutusyhtiöiltä tilastoaineistoa vuosilta 2002–2010 (ryhmäeläkevakuutuksen osalta vuosilta 2002–2009). Pyyntö koski Suomen liiketoimintaa. Aineisto sisältää seuraavat muuttujat:

- tilivuosi
- henkilön identifioiva tunnus, josta käy ilmi henkilön syntymäaika ja sukupuoli
- kuolinpäivämäärä tai kuolinvuosi, jos kuolinpäivämäärä ei ole saatavilla
- vakuutuksen voimaantulopäivämäärä. Ryhmäriskivakuutuksessa ja ryhmäeläkevakuutuksessa vakuutuksen voimaantulopäivämäärä tarkoittaa vakuutetun ryhmäänliittymispäivämäärää.
- vakuutuksen päättymispäivämäärä. Jos tiedossa oli vain vakuutuksen päättymisvuosi, päivämääräksi merkittiin päättymisvuoden viimeinen päivä. Ryhmäriskivakuutuksessa ja ryhmäeläkevakuutuksessa vakuutuksen päättymispäivämäärä tarkoittaa vakuutetun ryhmästäeroamispäivämäärää.
- korotusluokka vakuutuksen myöntöpäivänä (normaali henkivakuutus=0, korotettu henkivakuutus=1,2,3,5 tai 8)
- riskisumma henkivakuutuksessa ja vakuutussäästöjen määrä tai sen arvio eläkevakuutuksessa. Vuoden lopun tai kuoleman hetken summa. Riskisumma henkivakuutuksessa on kuolemanvarariskin varalta vakuutettu summa ilman lisävakuutuksia vähennettynä säästöillä ilman lisävakuutuksia.
- tariffikoodi
  - yksilöllinen säästövakuutus
  - yksilöllinen säästövakuutus unit linked (sisältää myös hybrideituotteet, joissa yhdistettyinä perustekorkoinen ja unit linked -vakuutus)
  - yksilöllinen riskivakuutus (pariturva käsitellään normaalina henkivakuutuksena)
  - ryhmäetuvakuutus

- ryhmäriskivakuutus
- yksilöllinen eläkevakuutus
- yksilöllinen eläkevakuutus unit linked (sisältää myös hybridituotteet, joissa yhdistettyinä perustekorkoinen ja unit linked -vakuutus)
- ryhmäeläkevakuutus (vain vanhuuseläkkeeseen liittyvä kuolinturva)
- perhe-eläke (voi rajoittaa koskemaan leskiä)
- tapaturmainen kuolema (tapaturmainen kuolema ja muu kuolema käsiteltiin erillisinä vakuutuksina)

# Liite B

# Väestökuolevuusennustetta vastaavat elinajanodotteet eri lähtövuosilla

| Nainen | | | | | |
|---|---|---|---|---|---|
| **0-vuotias** | | | | | |
| | **1878-2009** | **1924-2009** | **1960-2009** | **1955-2009** | **1989-2009** |
| **2010** | 83.3 | 83.3 | 83.3 | 83.3 | 83.3 |
| **2015** | 83.8 | 84.1 | 84.2 | 84.2 | 84.3 |
| **2020** | 84.4 | 84.8 | 85.1 | 85.1 | 85.2 |
| **2030** | 85.3 | 86.2 | 86.7 | 86.7 | 86.8 |
| **2050** | 87.0 | 88.4 | 89.5 | 89.5 | 89.5 |
| **30-vuotias** | | | | | |
| | **1878-2009** | **1924-2009** | **1960-2009** | **1955-2009** | **1989-2009** |
| **2010** | 53.8 | 53.9 | 53.9 | 53.9 | 53.9 |
| **2015** | 54.3 | 54.5 | 54.7 | 54.7 | 54.8 |
| **2020** | 54.8 | 55.2 | 55.5 | 55.5 | 55.6 |
| **2030** | 55.6 | 56.4 | 57.1 | 57.0 | 57.2 |
| **2050** | 57.2 | 58.4 | 59.7 | 59.7 | 59.8 |
| **50-vuotias** | | | | | |
| | **1878-2009** | **1924-2009** | **1960-2009** | **1955-2009** | **1989-2009** |
| **2010** | 34.6 | 34.7 | 34.7 | 34.7 | 34.7 |
| **2015** | 35.0 | 35.2 | 35.5 | 35.5 | 35.5 |
| **2020** | 35.4 | 35.8 | 36.2 | 36.2 | 36.3 |
| **2050** | 37.5 | 38.7 | 40.2 | 40.2 | 40.3 |
| **70-vuotias** | | | | | |
| | **1878-2009** | **1924-2009** | **1960-2009** | **1955-2009** | **1989-2009** |
| **2010** | 17.0 | 17.1 | 17.1 | 17.1 | 17.1 |
| **2015** | 17.3 | 17.5 | 17.7 | 17.7 | 17.9 |
| **2020** | 17.6 | 17.8 | 18.3 | 18.3 | 18.5 |
| **2030** | 18.1 | 18.6 | 19.4 | 19.4 | 19.8 |
| **2050** | 19.0 | 19.9 | 21.5 | 21.4 | 21.9 |

| Mies | | | | | |
|---|---|---|---|---|---|
| **0-vuotias** | | | | | |
| | **1878-2009** | **1924-2009** | **1960-2009** | **1955-2009** | **1989-2009** |
| **2010** | 76.6 | 76.7 | 76.7 | 76.7 | 76.8 |
| **2015** | 77.2 | 77.5 | 77.6 | 77.6 | 78.1 |
| **2020** | 77.7 | 78.3 | 78.6 | 78.6 | 79.3 |
| **2030** | 78.8 | 79.7 | 80.3 | 80.3 | 81.5 |
| **2050** | 80.6 | 82.0 | 83.4 | 83.4 | 85.4 |
| **30-vuotias** | | | | | |
| | **1878-2009** | **1924-2009** | **1960-2009** | **1955-2009** | **1989-2009** |
| **2010** | 47.6 | 47.6 | 47.7 | 47.7 | 47.7 |
| **2015** | 48.0 | 48.3 | 48.5 | 48.5 | 48.9 |
| **2020** | 48.5 | 48.9 | 49.4 | 49.4 | 50.1 |
| **2030** | 49.3 | 50.1 | 51.0 | 51.0 | 52.2 |
| **2050** | 50.9 | 52.2 | 54.0 | 53.9 | 55.8 |
| **50-vuotias** | | | | | |
| | **1878-2009** | **1924-2009** | **1960-2009** | **1955-2009** | **1989-2009** |
| **2010** | 29.2 | 29.3 | 29.3 | 29.3 | 29.4 |
| **2015** | 29.6 | 29.8 | 30.1 | 30.1 | 30.4 |
| **2020** | 29.9 | 30.3 | 30.9 | 30.8 | 31.4 |
| **2050** | 31.8 | 32.9 | 35.0 | 34.9 | 36.6 |
| **70-vuotias** | | | | | |
| | **1878-2009** | **1924-2009** | **1960-2009** | **1955-2009** | **1989-2009** |
| **2010** | 13.7 | 13.7 | 13.8 | 13.8 | 13.8 |
| **2015** | 13.9 | 14.0 | 14.3 | 14.3 | 14.6 |
| **2020** | 14.1 | 14.3 | 14.7 | 14.7 | 15.3 |
| **2030** | 14.6 | 14.8 | 15.6 | 15.6 | 16.6 |
| **2050** | 15.4 | 15.8 | 17.4 | 17.4 | 18.9 |

# Liite C

# Vertailut muiden tutkimuksien kanssa

| | TVL 1955-2008 | | | | | | K2012 LM 1955-2008 (tasoittamaton) | | | | | | EROTUS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| **Naiset 2010** | | | | | | | | | | | | | | | | | | |
| Lee-Miller | 83.4 | 58.8 | 34.7 | 21.3 | 13.1 | 6.5 | 83.4 | 58.8 | 34.7 | 21.3 | 13.1 | 6.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 82.9 | 58.4 | 34.3 | 21 | 12.8 | 6.4 | 82.9 | 58.4 | 34.3 | 21.0 | 12.8 | 6.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Ylä 80% | 83.8 | 59.2 | 35 | 21.6 | 13.3 | 6.6 | 83.8 | 59.2 | 35.0 | 21.6 | 13.3 | 6.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| **Naiset 2030** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| Lee-Miller | 86.7 | 62 | 37.6 | 23.8 | 15.1 | 7.6 | 86.7 | 62.0 | 37.6 | 23.8 | 15.0 | 7.6 | 0.0 | 0.0 | 0.0 | 0.0 | -0.1 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 85.3 | 60.6 | 36.3 | 22.7 | 14.2 | 7.1 | 85.3 | 60.6 | 36.3 | 22.7 | 14.2 | 7.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Ylä 80% | 88.1 | 63.3 | 38.8 | 24.9 | 15.9 | 8.2 | 88.0 | 63.2 | 38.8 | 24.8 | 15.8 | 8.1 | -0.1 | -0.1 | 0.0 | -0.1 | -0.1 | -0.1 |
| | | | | | | | | | | | | | | | | | | |
| **Naiset 2050** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| Lee-Miller | 89.6 | 64.8 | 40.2 | 26.1 | 16.9 | 8.8 | 89.5 | 64.6 | 40.1 | 26.0 | 16.8 | 8.7 | -0.1 | -0.2 | -0.1 | -0.1 | -0.1 | -0.1 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 87.6 | 62.8 | 38.4 | 24.5 | 15.6 | 8 | 87.6 | 62.8 | 38.4 | 24.5 | 15.6 | 7.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | -0.1 |
| Ylä 80% | 91.5 | 66.6 | 42 | 27.7 | 18.3 | 9.8 | 91.1 | 66.2 | 41.6 | 27.3 | 17.9 | 9.4 | -0.4 | -0.4 | -0.4 | -0.4 | -0.4 | -0.4 |
| | | | | | | | | | | | | | | | | | | |
| **Miehet 2010** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| Lee-Miller | 76.7 | 52.5 | 29.4 | 17.5 | 10.7 | 5.5 | 76.7 | 52.5 | 29.4 | 17.5 | 10.7 | 5.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 76.3 | 52.1 | 29 | 17.3 | 10.5 | 5.4 | 76.3 | 52.1 | 29.0 | 17.3 | 10.5 | 5.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Ylä 80% | 77.2 | 52.9 | 29.8 | 17.8 | 10.8 | 5.6 | 77.2 | 52.9 | 29.8 | 17.8 | 10.9 | 5.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| **Miehet 2030** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| Lee-Miller | 80.4 | 55.9 | 32.4 | 19.7 | 12.2 | 6.3 | 80.4 | 55.9 | 32.4 | 19.8 | 12.2 | 6.4 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 78.9 | 54.5 | 31.1 | 18.8 | 11.5 | 6 | 78.9 | 54.5 | 31.1 | 18.8 | 11.5 | 6.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Ylä 80% | 81.4 | 57.3 | 33.6 | 20.7 | 12.8 | 6.7 | 81.9 | 57.4 | 33.7 | 20.7 | 12.9 | 6.8 | 0.5 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 |
| | | | | | | | | | | | | | | | | | | |
| **Miehet 2050** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| Lee-Miller | 83.6 | 58.9 | 35.1 | 21.8 | 13.7 | 7.2 | 83.7 | 59.1 | 35.2 | 21.9 | 13.8 | 7.4 | 0.1 | 0.2 | 0.1 | 0.1 | 0.1 | 0.2 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 81.4 | 56.9 | 33.3 | 20.4 | 12.6 | 6.6 | 81.5 | 57.0 | 33.3 | 20.4 | 12.7 | 6.7 | 0.1 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 |
| Ylä 80% | 85.5 | 60.8 | 36.7 | 23.1 | 14.6 | 7.8 | 85.8 | 61.1 | 37.0 | 23.4 | 15.0 | 8.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.4 | 0.5 |

| | K2012 LM 1955-2009 (tasoittamaton) | | | | | | K2012 LM 1955-2009 (tasoitettu) | | | | | | EROTUS K2012 tasoittamaton-tasoitettu | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| **Naiset 2010** | | | | | | | | | | | | | | | | | | |
| Lee-Miller | 83.3 | 58.8 | 34.7 | 21.3 | 13.1 | 6.6 | 83.3 | 58.8 | 34.7 | 21.3 | 13.1 | 6.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 83.0 | 58.5 | 34.4 | 21.0 | 12.9 | 6.3 | | | | | | | | | | | | |
| Ylä 80% | 83.7 | 59.2 | 35.1 | 21.7 | 13.5 | 6.8 | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| **Naiset 2030** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| Lee-Miller | 86.7 | 62.0 | 37.6 | 23.9 | 15.1 | 7.7 | 86.7 | 62.0 | 37.6 | 23.9 | 15.1 | 7.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 85.4 | 60.9 | 36.6 | 22.9 | 14.3 | 7.2 | | | | | | | | | | | | |
| Ylä 80% | 87.8 | 63.0 | 38.7 | 24.8 | 15.9 | 8.1 | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| **Naiset 2050** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| Lee-Miller | 89.5 | 64.7 | 40.1 | 26.0 | 16.9 | 8.7 | 89.5 | 64.7 | 40.2 | 26.1 | 16.9 | 8.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 88.1 | 63.2 | 38.9 | 25.0 | 15.9 | 8.1 | | | | | | | | | | | | |
| Ylä 80% | 90.8 | 65.8 | 41.3 | 27.2 | 17.7 | 9.4 | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| **Miehet 2010** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| Lee-Miller | 76.7 | 52.4 | 29.3 | 17.3 | 10.5 | 5.4 | 76.7 | 52.4 | 29.3 | 17.3 | 10.5 | 5.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 76.3 | 52.1 | 29.0 | 17.0 | 10.3 | 5.2 | | | | | | | | | | | | |
| Ylä 80% | 77.1 | 52.8 | 29.7 | 17.7 | 10.8 | 5.7 | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| **Miehet 2030** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| Lee-Miller | 80.3 | 55.8 | 32.3 | 19.5 | 12.0 | 6.2 | 80.3 | 55.8 | 32.3 | 19.5 | 12.0 | 6.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 79.0 | 54.5 | 31.2 | 18.6 | 11.4 | 5.8 | | | | | | | | | | | | |
| Ylä 80% | 81.4 | 57.0 | 33.5 | 20.3 | 12.6 | 6.6 | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| **Miehet 2050** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** | **0** | **25** | **50** | **65** | **75** | **85** |
| Lee-Miller | 83.4 | 58.8 | 34.9 | 21.5 | 13.5 | 7.0 | 83.4 | 58.8 | 34.9 | 21.6 | 13.5 | 7.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | | | | | | | | |
| Ala 80% | 81.9 | 57.3 | 33.6 | 20.5 | 12.7 | 6.5 | | | | | | | | | | | | |
| Ylä 80% | 84.8 | 60.2 | 36.3 | 22.6 | 14.2 | 7.7 | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |

# Liite D

# $m_{x,t}$-, $p_{x,t}$- ja $e_x$-luvut

Erillisessä excel-tiedostossa ("K2012 raportti LiiteD") on esitetty luvussa 4.5 kuvatulla tavalla tehdyn kuolevuusennusteen $m_{x,t}$-taulut miehille ja naisille sekä niihin liittyvät hajonnat. Väestökuolevuuden luottamusvälin voi laskea luvussa 4.4 esitetyllä tavalla. Luvussa 5.6 on esitetty korjauskertoimet tuotteittain. Henkivakuutuskuolevuuden tietylle tuotteelle saa väestökuolevuuden $m_{x,t}$ ja tuotteen korjauskertoimen $L_x$ tulona. Henkivakuutuskuolevuutta laskettaessa tulee kuitenkin ottaa huomioon, että tässä raportissa korjauskertoimet on esitetty vain tietylle ikävälille ottamatta kantaa ikävälin ulkopuolella olevien korjauskertoimien arvoihin.

Tarja Siréniltä (tarja.siren@fiva.fi) on saatavilla excel-taulukko erikseen miehille ja naisille, joissa on esitetty havaittu väestökuolevuus vuosilta 1955–2009 (tasoitettu HMD-tietokannan aineisto) sekä väestökuolevuusennuste vuosille 2010–2129. Taulukossa on esitetty myös Lee-Miller -mallin parametrit $a_x$, $b_x$ ja $k_t$ sekä väestökuolevuusennusteeseen liittyvä hajonta ja luottamusvälit. Lisäksi taulukossa on esitetty luvun 5.6 mukaiset korjauskertoimet sekä niihin perustuvat henkivakuutuskuolevuudet luottamusväleineen (luottamusväleissä on otettu huomioon ainoastaan väestöenusteeseen liittyvä epävarmuus). Laskentataulukossa on mahdollista muuttaa käytettyjä korjauskertoimia. Henkivakuutetuille on laskettu yhden vuoden selviämistodennäköisyydet ennustejaksolla aktuaariestimaattorilla

$$p_x = \frac{2 - m_x}{2 + m_x}.$$

ja elinajanodotteet kaavalla:

$$e_x = \frac{1}{2} + \sum_{k=1}^{\infty} {}_k p_x,$$

missä

$${}_k p_x = \prod_{u=0}^{k-1} p_{x+u}.$$

Excelissä on muodostettu myös henkivakuutuskuolevuutta vastaava kohorttiaineisto luottamusväleineen. Näissä luottamusväleissä on suositeltavaa käyttää vakioitua parametrin $b_x$ arvoa.

# Liite E

# Henkivakuutuskuolevuuden epävarmuuden huomioon ottaminen

Tässä liitteessä esitellään lyhyesti, miten henkivakuutuskuolevuuden epävarmuus voidaan ottaa huomioon. Henkivakuutuskuolevuus on muotoa

$$\hat{m}^H_{x,t} = \hat{L}_x \hat{m}_{x,t}.$$

Jos korjauskerroin $\hat{L}_x$ ja väestökuolevuus $\hat{m}_{x,t}$ oletetaan riippumattomiksi, epävarmuutta voidaan arvioida esimerkiksi normaalijakaumaoletukseen perustuvilla luottamusväleillä. Koska

$$\ln(\hat{m}^H_{x,t}) = \ln(\hat{L}_x) + \ln(\hat{m}_{x,t}),$$

niin riippumattomuudesta seuraa, että varianssi on

$$\mathrm{Var}(\ln(\hat{L}_x) + \ln(\hat{m}_{x,t})) = \mathrm{Var}(\ln(\hat{L}_x)) + \mathrm{Var}(\ln(\hat{m}_{x,t})).$$

Näin ollen luottamusvälit ovat muotoa

$$\ln(\hat{m}^H_{x,t+h}) \pm z_{\alpha/2}\sqrt{\hat{\sigma}^2_{\ln(\hat{m}_{x,t})} + \hat{\sigma}^2_{\ln(\hat{L}_x)}}$$

ja

$$\hat{m}^H_{x,t+h}\exp(\pm z_{\alpha/2}\sqrt{\hat{\sigma}^2_{\ln(\hat{m}_{x,t})} + \hat{\sigma}^2_{\ln(\hat{L}_x)}})$$

(vertaa luvut 4.4 ja 5.6).

# Liite F

# Summakuolevuuksien sovitteet

Tässä liitteessä esitellään summakuolevuuden perusteella määritellyt korjauskertoimet yksilölliselle riskivakuutukselle sekä ryhmäriski- ja ryhmäetuvakuutuksille. Korjauskertoimet on määritelty sovittamalla regressiokäyrä summakuolevuuden ja väestökuolevuuden suhteeseen ($\frac{m_x^S}{m_x}$, missä $m_x$ on väestökuolevuuksien keskiarvo vuosilta 2002–2009). Regressioanalyysi ei ole kaikissa tapauksissa optimaalisin menetelmä korjauskertoimen määrittämiseen siihen liittyvien oletusten takia (muun muassa vakiovarianssi). Lisäksi tuloksia analysoitaessa tulee ottaa huomioon volyymisuureaineistoon liittyvät puutteet ja epävarmuus kuten yhden yhtiön tietojen puuttuminen volyymiaineistosta.

Yksilöllisellä riskivakuutuksella summakuolevuuksien perusteella määritellyt korjauskertoimet ovat seuraavat:

$$L_x^{\mathrm{S,\ miehet}} = -0.02930 + 0.00907x$$

$$L_x^{\mathrm{S,\ naiset}} = 0.21033 + 0.00995x$$

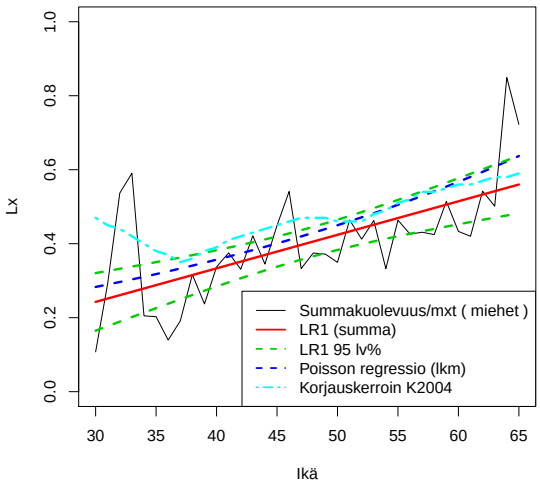
Kuvassa F.1 on esitetty miesten K2004-korjauskerroin, Poisson-regressiolla määritelty korjauskerroin sekä summakuolevuuteen perustuva korjauskerroin luottamusväleineen. Huomataan, että summakuolevuuden perusteella määritelty korjauskerroin on hiukan Poisson-regressiolla määriteltyä korjauskerrointa alempana. Naisilla taas Poisson-regressiolla määritelty korjauskerroin on alemmalla tasolla kuin summakuolevuuteen perustuva korjauskerroin (kuva F.2). Naisilla regressiosuoran tasoon vaikuttaa muutama poikkeavan suuri havainto.

Ryhmäriski- ja ryhmäetuvakuutusten summakuolevuuteen perustuvat korjauskertoimet ovat seuraavat:

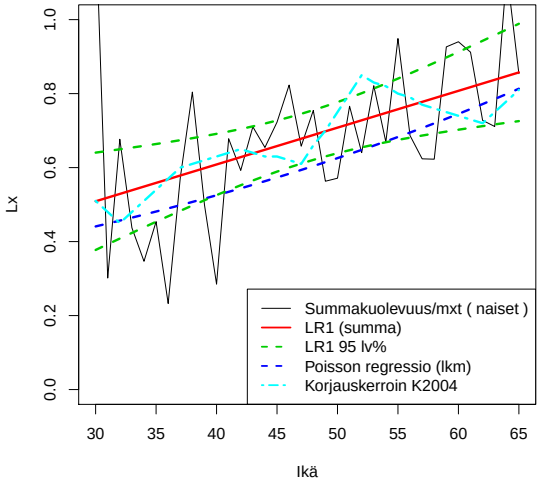
$$L_x^{\mathrm{S,\ miehet}} = -0.01627 + 0.00615x$$

$$L_x^{\mathrm{S,\ naiset}} = 0.17464 + 0.00556x$$

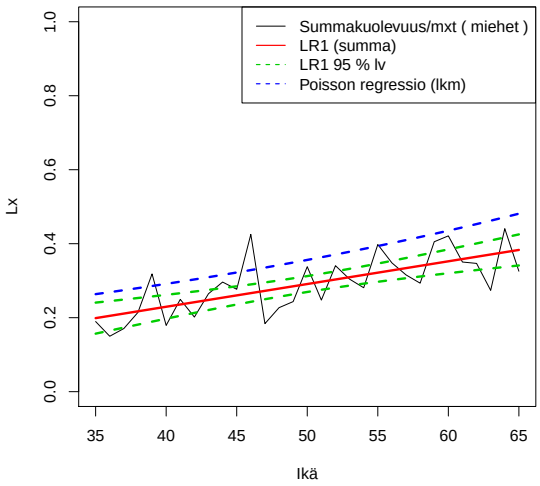
Naisilla ja miehillä Poisson-regressioon perustuvat korjauskertoimet ovat korkeammalla tasolla kuin summakuolevuuteen perustuvat korjauskertoimet. Kun otetaan huomioon korjauskertoimiin liittyvä epävarmuus, eivät erot ole merkittäviä.



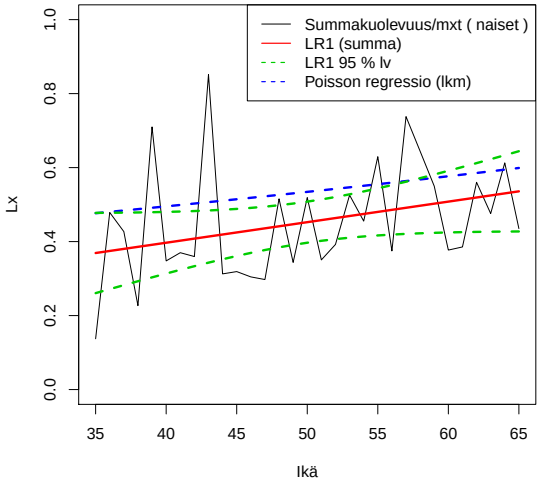

Kuva F.1: Summakuolevuuden ja väestökuolevuuden suhteeseen sovitettu suora (LR1) regressioanalyysillä. Lisäksi 95 %:n luottamusväli LR1:lle (LR1 95 % lv) sekä Poisson regressioon perustuva korjauskerroin. (Yksilöllinen riskivakuutus, miehet)




Kuva F.2: Summakuolevuuden ja väestökuolevuuden suhteeseen sovitettu suora (LR1) regressioanalyysillä. Lisäksi 95 %:n luottamusväli LR1:lle (LR1 95 % lv) sekä Poisson regressioon perustuva korjauskerroin. (Yksilöllinen riskivakuutus, naiset)




Kuva F.3: Summakuolevuuden ja väestökuolevuuden suhteeseen sovitettu suora (LR1) regressioanalyysillä. Lisäksi 95 %:n luottamusväli LR1:lle (LR1 95 % lv) sekä Poisson regressioon perustuva korjauskerroin. (Ryhmäetu- ja ryhmäriskivakuutukset, miehet)

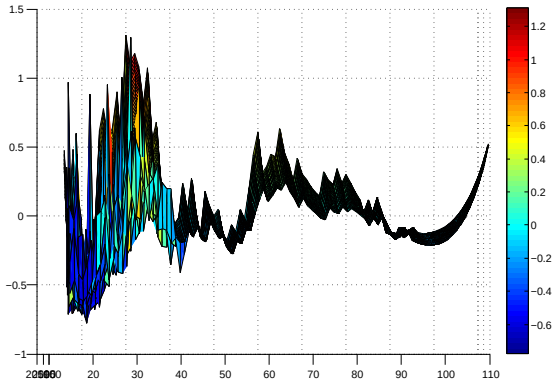



Kuva F.4: Summakuolevuuden ja väestökuolevuuden suhteeseen sovitettu suora (LR1) regressioanalyysillä. Lisäksi 95 %:n luottamusväli LR1:lle (LR1 95 % lv) sekä Poisson regressioon perustuva korjauskerroin. (Ryhmäetu- ja ryhmäriskivakuutukset, naiset)

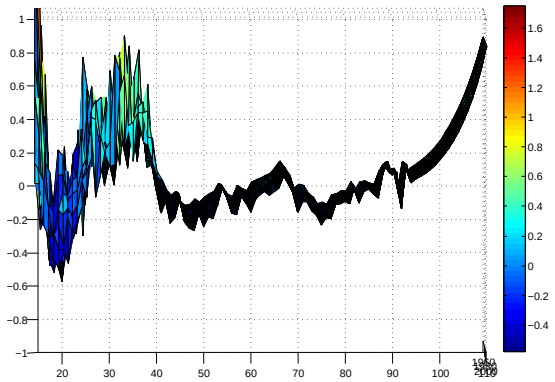
# Liite G

# K2004-tutkimusta vastaavan henkivakuutuskuolevuuden ja sovitteen suhteellinen ero

Tässä liitteessä on havainnollistettu kuvaajien avulla K2004-korjauskertoimilla korjatun Lee-Carter -ennusteen (lähtövuodet 1960–2001) sekä K2004-sovitteen suhteellista eroa ($(m_{x,t}^{LC} - m_{x,t}^{K2004})/m_{x,t}^{K2004}$).




Kuva G.1: K2004-sovitteen ja henkivakuutuskuolevuuden suhteellinen ero kohorteille 1960-2011 (naiset).




Kuva G.2: K2004-sovitteen ja yksilöllisen henkivakuutuskuolevuuden suhteellinen ero kohorteille 1960-2011 (miehet).

# Liite H

# R-ohjelmiston ohjeet

Ohjetta tehtäessä käytössä on ollut R-ohjelmiston versio 2.15.0. Rob J. Hyndmanin R-paketista demography on ollut käytössä versio 1.12 (`http://robjhyndman.com/software/demography/`). Lisäksi demography-paketti vaatii muun muassa seuraavia paketteja forecast, rainbow, ftsa. Myös R-paketti MASS on hyödyllinen.

## H.1 Tavoite

Tämän ohjeen tarkoituksena on antaa valmiudet toistaa K2012-tutkimuksen laskelmat kuolevuusennusteen osalta. Lisäksi ohjeessa annetaan valmiudet Poisson-regression suorittamiseen R-ohjelmistolla. Tarkemmat ohjeet ja määritelmät Demography-pakettiin liittyen löytyy Rob J. Hyndmanin kotisivuilta lataustiedostojen yhteydestä.

Ohjeessa ei käydä läpi R-ohjelmiston perusteita. Jos ohjelmisto on uusi, kannattaa tutustua muun muassa Crawley (2007) kirjaan sekä R-ohjelmiston kotisivuilta löytyvään materiaaliin (`www.r-project.org`). Lisäksi Internetissä on saatavilla paljon kurssimateriaalia R-ohjelmistoon liittyen.

Alle on listattu joitain R-ohjelmiston peruskomentoja. Kokeilemalla saa käsityksen, miten komennot toimivat. Lisäksi funktioiden toiminnallisuuksista saa lisätietoa R-ohjelmiston Help-valikosta tai antamalla käskyn "?funktion nimi".

```
?glm

# kommentteja

1:10

c(1,7,9)

rep(0,5)

seq(0,10, length.out=2)

set.seed(10)

a <- matrix(rnorm(100), nrow=20)

a[1 : 2, 5]
```

?tapply

Lisäksi alla on muutama yleisemmistä kuvaajatyypeistä:

?plot

?lines

?hist

?barplot

?boxplot

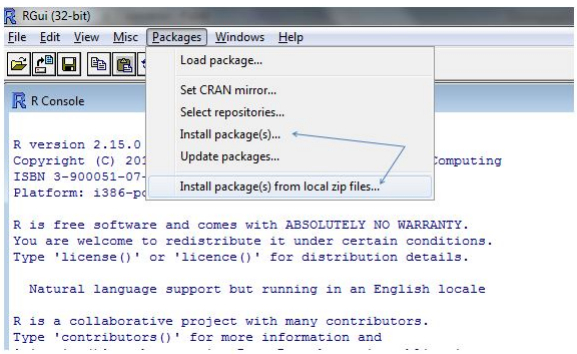
Käskyllä ”par(mfrow=c(r,s))” saa useita kuvaajia yhteen näkymään. Rivien lukumäärään viittaa r ja sarakkeiden lukumäärään s. For-silmukan avulla voi latoa useita kuvaajia samaan kuvaajaan. Tällöin käskyt ”par(new=TRUE)” ja ”par(new=FALSE)” ovat hyödyllisiä.

Lopuksi vielä muutama sana taulukoiden tallentamisesta. R-taulukoita voi tallettaa muun muassa txt-tiedostoiksi. Alla olevassa koodiesimerkissä ”ex” viittaa R-ohjelmassa olevaan tauluun nimeltä ”ex”. Tämän jälkeen määritellään polku, jonne aineisto halutaan tallentaa. Käsky ”sep” kertoo erottimen.

write.table(ex,"C:/... /taulukot /ex_F.txt", sep=” ”)

## H.2 R-ohjelman alustus

Ensimmäisenä tulee tallentaa tarvittavat paketit. Yksi tapa on tallentaa paketit zip-paketteina omalle koneelle ja ladata ne R:n kohdasta ”Install package(s) from local zip files” (paketit on saatavilla zip-tiedostoina muun muassa demography-paketin kotisivuilta). Toinen tapa on valita kohta ”Install package(s)” ja valita maantieteellisesti lähin maa, josta paketin lataa (kuva H.1). Tämän jälkeen pakettien lataaminen R:n käyttöön onnistuu ”library”-




Kuva H.1: Pakettien lataus R-ohjelmistossa

käskyllä seuraavasti:

library(ftsa)

```
library(forecast)

library(demography)
```

Aineistojen ja kuvaajien tallennusta varten voi asettaa hakemiston, jonne kuvat tallennetaan.

```
getwd() #kertoo nykyisen hakemiston

setwd("C:/kuvat") # asettaa uuden hakemistopolun
```

Tärkeä työkalu valmisfunktioiden käytössä on tietää, kuinka funktiot toimivat. Koodin saa näkyviin kirjoittamalla funktion nimen R:ään, esim. "lca", tai, jos tämä ei toimi, "demography:::fitmx" (demography kertoo paketin, jossa funktiota on käytetty, ja fitmx funktion nimen).

## H.3 Kuolevuus

### H.3.1 Aineiston lataaminen ja valinta

Suomen väestöaineisto on ladattu Human Mortality Databasesta (`www.mortality.org`). Aineiston lataaminen vaatii kirjautumista. Demography paketti vaatii, että aineisto ladataan R-ohjelmaan tietyssä muodossa. Tasoittamattoman aineiston voi ladata suoraan HMD-tietokannasta käyttäjän omilla tunnuksilla funktiolla "hmd.mx":

```
Suomi <- hmd.mx("FIN", "käyttäjänimesi", "salasanasi", "Finland")
```

Vastaavan aineiston voi myös ladata txt-tiedostosta tallentamalla HMD:stä Suomen aineiston omalle tietokoneelle. HMD:ssa aineistot on nimetty seuraavasti:

"Exposure to Risk 1x1" (Finland_exp) ja "Death Rates 1x1" (Finland_mx).

Seuraavassa koodissa on oletettu, että aineistot on tallennettu käyttäjän työasemalla (Working Directory):

```
fin<-read.demogdata("Finland_mx.txt", "Finland_exp.txt",
    type="mortality", label="Finland")
```

HMD-tietokannassa on myös saatavilla aineisto, jossa mxt-luvut sisältävät tasoituksen ("Life tables"). Tasoituksen määrittely on nähtävillä HMD-tietokannan protokollassa (`www.mortality.org/Public/Docs/MethodsProtocol.pdf`). K2012-raportin lopullisessa kuolevuusennusteessa on käytetty HMD-tietokannan tasoitettua Suomen aineistoa.

Suomen demodataa "fin" voidaan nyt hyödyntää monissa demography-paketin funktiossa. Huomaa kuitenkin, että käskyllä "fin" saa vain tiedot demodatan sisällöstä. Esimerkiksi ikävektorin tai vuosivektorin saa käskyillä "fin$age" ja "fin$year". Käsky "fin$rate" antaa keskikuolevuudet naisille, miehille ja yhteensä. Tässä muodossa aineistoa ei voi kuitenkaan hyödyntää

esim. kuvaajia tehtäessä. Tästä syystä joskus on tarvetta tutkia pelkästään esim. naisten keskikuolevuuksia, mikä onnistuu käskyllä ”fin$rate$female”.

Demography-paketissa on valmiina funktioita, joilla voi tarvittaessa rajoittaa aineiston ikäväliä tai tarkasteltavia vuosia. Funktio ”extract.years” rajaa tarkasteltavat vuodet ja ”extract.ages” tarkasteltavat iät. Käytettäessä funktiota ”extract.ages” tulee olla tarkkana, että koko todennäköisyysmassa tulee otettua huomioon. Tästä pitää huolen valinta ”combine.upper =TRUE”.

```
fin_a <- extract.ages(fin, ages=0:100, combine.upper=TRUE)

fin_y<- extract.years(fin, 1955:2009)
```

### H.3.2 Kuolevuusennuste

Kuolevuusennusteen saa tehtyä kahdessa vaiheessa. Ensimmäisessä vaiheessa tehdään estimointi havaitusta aineistosta ja toisessa vaiheessa ennuste.

Estimointi tapahtuu funktiolla ”lca”. Funktiolle syötearvoina annetaan data (fin), tarkasteltavat vuodet (years), sukupuoli (series), maksimi-ikä (max.age) ja menetelmä (”none” vastaa K2004:ssä käytettyä menetelmää, ”dt” Lee-Carterin alkuperäistä menetelmää ja ”e0” Lee-Miller -menetelmää eli tässä raportissa käytettyä menetelmää (vrt. luku 4)) Lisäksi annetaan totuusarvo, estimoidaanko lähtöaineiston nolla-arvoiset keskikuolevuudet viereisten ikäryhmien kuolevuuksilla (interpolate). Alla on esitetty eri menetelmiin liittyvät koodit. K2004-referenssikuolevuutta vastaava koodi on muotoa:

```
LC <- lca(fin, years=1955-2009, series= ”female”, max.age=100,
    adjust=”none”, interpolate=TRUE)
```

Leen ja Carterin alkuperäisessä paperissa esitetty menetelmä:

```
LCdt <- lca(fin, years=1955-2009, series= ”female”, max.age=100,
    adjust=”dt”, interpolate=TRUE)
```

Lee-Miller (K2012):

```
LM <- lca(fin, years=1955-2009, series= ”female”, max.age=100,
    adjust=”e0”, interpolate=TRUE)
```

Huomaa, että jos series-valinta jätetään ottamatta huomioon, laskenta tehdään naisille. Funktio lca antaa tuloksenaan muun muassa mallin parametrit ax, bx, kt (esim. ax saadaan käskyllä LM$ax) sekä aikasarjamallin residuaalit ja sovitetut arvot.

Demography-paketissa on mahdollisuus tehdä estimointi muun muassa Booth-Maindonald-Smith -menetelmällä funktiolla ”bms”. Esimerkki funktion käytöstä sekä syötearvojen selitykset löytyvät demography-paketin virallisesta protokollasta.

Kuolevuusennuste h vuodelle tehdään funktiolla ”forecast.lca”. Menetelmä on vastaava kuin luvussa 4.2 on mainittu eli ”random walk with drift”. Huomaa, että Lee-Carter -menetelmässä ”jumpchoice” on ”fit” eli estimoidut arvot, kun taas Lee-Millerissä ”actual” eli havaitut arvot. Termi ”se” viittaa hajontatermiin. Näin ollen sillä on vaikutusta luottamusväleihin. Mahdolliset

vaihtoehdot ovat ”innovonly” ja ”innovdrift”. ”Innovonly” ottaa huomioon ainoastaan menetelmään liittyvän epävarmuuden, kun taas ”innovdrift” ottaa huomioon menetelmän epävarmuuden lisäksi drift-termiin liittyvän epävarmuuden (vrt. luku 4.4). Lisäksi on mahdollisuus antaa luottamusvälien taso (”level”). Oletusarvoisesti funktio muodostaa 80 %:n luottamusvälin.

K2004-referenssikuolevuutta vastaava ennuste on seuraava:

```
LC_f <- forecast.lca(lca,h=50, jumpchoice="fit", se="innovdrift", level=95)
```

Leen ja Carterin alkuperäisessä artikkelissa esitetty menetelmä on seuraava:

```
LCdt_f <- forecast.lca(LCdt, h=50, jumpchoice="fit", se="innovdrift", level=95)
```

Lee-Miller -menetelmä on seuraava:

```
LM_f <- forecast.lca(LM, h=100, jumpchoice="actual", se="innovdrift", level=95)
```

Funktio forecast.lca antaa tuloksenaan muun muassa ennustetun kuolevuustaulukon, joka sisältää ennusteen lisäksi luottamusvälin ala- ja ylärajan (käsky ”LC_f$rate”). Tässäkin tapauksessa pelkän ennusteen saa käskyllä LM_f$rate$female (toisin sanoen ”female” viittaa aikaisemmin valittuun sukupuoleen (”series”)) ja esimerkiksi luottamusvälin ylärajan käskyllä ” LM_f$rate$upper”. Jos ennusteen yläraja halutaan vuodelta 1955 0–9-vuotiaille, saadaan se käskyllä ”LM_f$rate$upper[1:10,1]”. Matriisin riveillä on iät ja sarakkeilla vuodet.

### H.3.3 Elinajanodotteiden laskenta

Elinajanodotteet saadaan laskettua funktioilla ”life.expectancy” ja ”flife.expectancy”. Funktiot eroavat toisistaan sillä, että ”flife.expectancy” laskee myös luottamusvälin elinajanodotteelle. Funktio ”flife.expectancy” ei ota kuitenkaan huomioon $k_t$ drift-termiin liittyvää hajontaa kokonaan, mistä syystä luottamusväleistä muodostuu liian kapeita. Näin ollen funktion käyttöä kannattaa välttää. Funktion koodin korjaaminen onnistuu melko helposti.

Funktio ”life.expectancy” ei laske luottamusvälejä, mutta luottamusvälejä voi arvioida syöttämällä funktiolle ennustettujen keskikuolevuuksien ala- tai ylä-rajan. Tämä on parempi tapa kuin korjaamattoman ”flife.expectancy” -funktion käyttö. Alla esimerkki ”life-expectancy”-funktion käytöstä:

```
life.expectancy(LM_f, age=0, series=sex, type="period", max.age=100)
```

## H.4 Poisson-regressio

Poisson-regression saa suoritettua R-ohjelmistossa funktiolla ”glm”. Alla on luvun 5.6 mallia 3 vastaava esimerkki. Toisin sanoen malli on muotoa $\ln(L_x) = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 sx$. Selitettävänä muuttujana on kuolleiden lukumäärä. Selittäviksi muuttujiksi on valittu ikä ja sukupuoli (mallin voisi tehdä myös

erikseen miehille ja naisille). Mallissa offset-muuttujana on odotettu kuolleiden lukumäärä eli ($M_x E_x$). Lisätietoja mallista ja aineistosta voi lukea luvusta 5.

attach(data) # liitetään aineiston ”data” muuttujat työtilaan. Näin jatkossa ei tarvitse määrittää erikseen, mistä aineistosta on kyse.

model <- glm(Lkm_kuo ∼ ika*sex + offset(log(expo))+ log(mxt)), family=poisson(link=log))

Mallin yhteenvedon saa komennolla

summary(model, correlation=TRUE).

Komento ”correlation=TRUE” on hyvä lisätä tilanteissa, joissa halutaan tutkia myös parametrien korrelaatio- ja/tai kovarianssimatriisia. Kovarianssimatriisin saa muun muassa komennolla

summary.glm(model, correlation=TRUE)$cov.unscaled.

Mallin kertoimet saa komennolla ”model$coefficients”. Esimerkiksi naisten suoran saa määriteltyä seuraavasti

x <- 35:70 # ikäväli

lnLx_N <- model$coefficients[1] + model$coefficients[2] * x.