

PRO GRADU -TUTKIELMA

Katja Laine

**Aleksitymian ja kuolleisuuden välisen yhteyden
selvittäminen yleistettyjä additiivisia malleja käyttäen**

TAMPEREEN YLIOPISTO
Informaatiotieteiden yksikkö
Tilastotiede
Marraskuu 2012

Tampereen yliopisto

Informaatiotieteiden yksikkö

LAINEN, KATJA: Aleksitymian ja kuolleisuuden välisen yhteyden selvittäminen yleistettyjä additiivisia malleja käyttäen

Pro gradu -tutkielma, 38 s., 10 liites.

Tilastotiede

Marraskuu 2012

Tiivistelmä

Aleksitymiällä tarkoitetaan persoonallisuuden piirrettä, jolle on ominaista tunteiden tunnistamisen ja kuvailemisen vaikeus sekä ulkoistava ajattelutapa. Aleksityymisyyttä tutkitaan 20-osioisen Toronton aleksitymiakyselyn (TAS-20) avulla. Jos vastaajan kyselystä saama pistemäärä on vähintään 61 pistettä, vastaaja määritellään aleksityymiseksi. Aleksitymia on yhteydessä moniin sairauksiin, mielenterveyden häiriöihin ja epäterveellisiin elämäntapoihin.

Tutkielman tavoitteena on selvittää, onko aleksitymiällä yhteyttä kuolleisuuteen. Analysointimenetelmänä käytetään yleistettyjä additiivisia malleja. Tutkimus perustuu Terveys 2000 -aineistoon, joka kerättiin Suomessa vuosina 2000–2001. Tutkielmassa on käytetty vuoden 2008 loppuun mennessä kertyneitä kuolleisuustietoja. Aleksitymian yhteyttä tarkastellaan kokonaiskuolleisuuteen ja kolmeen eri kuolinsyyhyyn, jotka ovat sydän- ja verisuonitaudit, verenpaine- tai iskeemiset sydänsairaudet ja aivoinfarkti-, verenpaine- tai iskeemiset sydänsairaudet. Lisäksi aleksitymian mahdollista yhdysvaikutusta iän, koulutuksen ja painoindeksin suhteen tutkittiin interaktiomallien avulla.

Aleksitymiällä ei havaittu yhteyttä kokonaiskuolleisuuteen tavallisessa mallissa, mutta interaktiitermien yhteys kuolleisuuteen oli positiivinen ja aleksityymisyyden omavaikutuksen yhteys negatiivinen. Kun kuolemia rajattiin eri kuolinsyyden mukaan, aleksitymiällä havaittiin lievä yhteys kuolleisuuteen tavallisissa malleissa. Sydän- ja verisuonitautikuolleisuuden interaktiomallissa aleksityymisyyden omavaikutus ei ollut tilastollisesti merkitsevä, mutta kaikilla interaktiotermeillä oli merkitsevä positiivinen yhteys kuolleisuuteen. Verenpaine- tai sydänsairauskuolleisuuden interaktiomallissa vain iän ja aleksityymisyyden välisellä interaktiolla oli positiivinen yhteys alle 50- ja yli 75-vuotiaiden kuolleisuuteen. Aivoinfarkti-, verenpaine- tai sydänsairauskuolemien tarkastelussa interaktiomalli voimisti aleksitymian omavaikutuksen positiivista yhteyttä kuolleisuuteen, mutta ikää lukuun ottamatta interaktiitermien vaikutus oli negatiivinen.

Asiasanat Aleksitymia, kuolleisuus, yleistetty additiivinen malli, logistinen regressio, splini

Sisältö

1 Johdanto	4
1.1 Johdatus aleksitymiaan	4
1.2 Tutkielman tavoite ja rakenne	4
2 Aineisto	6
2.1 Aineiston kerääminen	6
2.2 Aleksitymia	8
2.3 Kuolleisuus	9
2.4 Taustamuuttujat	10
2.4.1 Kvantitatiiviset muuttujat	10
2.4.2 Kvalitatiiviset muuttujat	12
3 Tutkimusmenetelmät	14
3.1 Yleistetyt lineaariset mallit (GLM)	14
3.2 Yleistetyt additiiviset mallit (GAM)	15
3.3 Tasoitusmenetelmät	17
3.3.1 Johdatus splineihin	17
3.3.2 Thin plate -regressiosplinit	19
3.3.3 Tensoritulotasoite	19
4 GAM:in soveltaminen aineistoon	22
4.1 Mallin sovittaminen	22
4.2 Tulokset	23
5 Yhteenveto	35
Lähteet	37
Liite A: TAS-20-kyselylomake	39
Liite B: Mallien tuloksia	40
Liite C: Mallien vertailua	48

1 Johdanto

1.1 Johdatus aleksitymiaan

Aleksitymiaksi kutsutaan persoonallisuuden piirrettä, jolle on ominaista tunteiden tunnistamisen ja kuvailemisen vaikeus sekä ulkoistava ajattelutapa. Aiemmissä tutkimuksissa aleksitymian on todettu olevan yhteydessä useisiin somaattisiin sairauksiin ja mielenterveyden häiriöihin, kuten esimerkiksi korkeaan verenpaineeseen, liikalihavuuteen, astmaan, diabetekseen, krooniseen kipuun, masennukseen ja somatisaatiohäiriöön. Aleksitymia on liitetty myös alkoholin liikkakäyttöön, päihteiden käyttöön, liikunnan puutteeseen ja epäterveellisiin ruokailutottumuksiin. (Chatzi et al. 2009; Helmers & Mente 1999; Lumley, Stettner & Wehmer 1996; Lumley, Asselin & Norman 1997; Marchesi, Brusamonti, & Maggini 2000; Serrano et al. 2006.) Lisäksi Mattila (2009) on osoittanut väitöskirjassaan, että aleksitymiolla on yhteys myös muun muassa alhaiseen koulutus- ja tulotasoon, naimattomuuteen tai leskeyteen sekä työuupumukseen. Ei ole kuitenkaan selvää, johtaako aleksityymisyys tiettyihin terveysongelmiin ja elämäntapoihin vai altistavatko huonot elintavat ja terveysongelmat aleksitymian kehittymiselle. Yhteyksien syy-seuraussuhteet vaativat vielä lisätutkimusta, mutta erilaisia hypoteesejakin on kuitenkin esitetty. (Tolmunen, Lehto, Heliste, Kurl & Kauhanen 2010; Lumley, Stettner & Wehmer 1996.)

Aleksitymian esiintymistä kartoitetaan kyselylomakkeella. Yleisimmin Suomessa käytetään 20-osioista Toronton aleksitymiakyselyä (Toronto Alexithymia Scale, TAS-20) (liite A). Kysymykset ovat Likert-asteikollisia, ja vastaukset pisteytetään siten, että ”täysin eri mieltä” -vaihtoehdosta saa yhden ja ”täysin samaa mieltä” -vaihtoehdosta viisi pistettä. Kun eri kysymysten pisteet lasketaan yhteen, saadaan TAS-summa, jonka perusteella voidaan arvioida vastaajan aleksityymisyyttä. Kysymysten 4, 5, 10, 18 ja 19 asteikot käännetään ennen yhteenlaskua. Katkaisupiste aleksityymisyydelle on 61 pistettä, eli ne, jotka saavat kyselystä vähintään 61 pistettä, määritellään vahvasti aleksityymisiksi. Suomalaisista 9,9 % on vahvasti aleksityymisiä, ja se on miehillä (11,9 %) yleisempää kuin naisilla (8,1 %) (Mattila 2009).

1.2 Tutkielman tavoite ja rakenne

Tutkielman tavoitteena on selvittää aleksitymian yhteyttä kuolleisuuteen Suomen väestössä. Tutkimus perustuu Kansanterveyslaitoksen organisoimaan Terveys 2000 -aineistoon, joka kerättiin Suomessa vuosina 2000–2001 kaksivaihei-

sella ryväsotannalla. Analysointiin käytetään yleistettyjä additiivisia malleja (GAM-mallit), joiden etuna yleistettyihin lineaarisiin malleihin (GLM) nähden on joustavuus monimutkaisten riippuvuuksien kuvaamisessa. Analyysit suoritetaan R-ohjelmiston mgcv-kirjastosta löytyvällä gam-funktiolla.

Aiheesta ei ole aiemmin tehty kattavaa väestötason tutkimusta, mutta pienemmällä tutkimusjoukoilla aleksitymian on todettu olevan yhteydessä kohonneeseen kuolleisuuteen (Kauhanen, Kaplan, Cohen, Julkunen & Salonen 1996; Kojima et al. 2010; Tolmunen, Lehto, Heliste, Kurl & Kauhanen 2010). Aiempien tutkimusten mukaan aleksitymia on itsenäinen kuolleisuuden riskitekijä keski-ikäisillä miehillä (Kauhanen et al. 1996) ja hemodialyysipotilailla (Kojima et al. 2010). Kauhasen et al. (1996) mukaan aleksitymia on yhteydessä etenkin onnettomuuksista, itsemurhista ja henkirikoksista aiheutuneeseen kohonneeseen kuolleisuuteen ja Tolmuseen et al. (2010) mukaan sydän- ja verisuonitautikuolleisuuteen keski-ikäisillä miehillä.

Tutkielma rakentuu viidestä luvusta. Johdannon jälkeen toisessa luvussa esitellään tarkemmin aineisto, sen keräystapa ja keskeiset muuttujat. Kolmannessa luvussa tarkastellaan yleistettyjä lineaarisia malleja koskevaa menetelmäteoriaa, yleistettyjä additiivisia malleja sekä erilaisia tasoitusmenetelmiä. Neljäs luku yhdistää kahden aiemman luvun asiat, kun GAM-malleja sovitaan aineistoon. Lopuksi viidennessä luvussa tehdään yhteenveto tutkimuksen onnistumisesta ja saaduista tuloksista.

2 Aineisto

2.1 Aineiston kerääminen

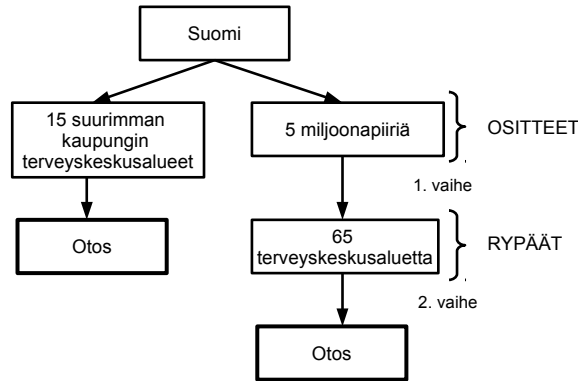
Terveys 2000 -aineisto kerättiin Suomessa vuosien 2000–2001 aikana. Kansanterveyslaitoksen hallinnoimaan hankkeeseen osallistui suuri joukko erilaisia sosiaali- ja terveysalan organisaatioita sekä esimerkiksi Kansaneläkelaitos, Stakes, Tilastokeskus ja UKK-instituutti. Kohdejoukkona oli Suomen 18 vuotta täyttänyt väestö. Tutkittaville tehtiin terveyshaastattelu ja 30 vuotta täyttäneille lisäksi terveystarkastus. (THL 2009.)

Otoksen haluttiin olevan kansallisesti edustava ja kustannustehokkaasti poimittu. Siksi poiminta suoritettiin kaksivaiheisena ositettua ryväotantaa hyödyntäen. Otokseen poimittiin kohdehenkilöitä yhteensä 80 eri terveyskeskusalueelta, jotka voivat muodostua joko yksittäisistä kunnista tai useamman kunnan muodostamista kuntayhtymistä. Poimintaa varten Suomi jaettiin viiteen alueeseen, jotka vastaavat likimain yliopistosairaalapiirejä. Näitä piirejä kutsutaan miljoonapiireiksi, koska kunkin yliopistosairaalapiirin alueella on likimain miljoona asukasta. Osituksessa käytettiin väestömäärään suhteutettua suhteellista kiintiöintiä. Ahvenanmaata ja saaristoa ei otettu mukaan tutkimukseen huonon saavutettavuuden vuoksi. Viiden ositteen sisällä oli yhteensä 249 terveyskeskuspiiriä eli ryvästä. (Heistaro 2005.)

Viidentoista suurimman kaupungin haluttiin tulevan otokseen automaattisesti. Loput 65 terveyskeskuspiiriä poimittiin viidestä miljoonapiiristä siten, että suuret kaupungit mukaan lukien kustakin poimittiin 16 terveyskeskusalueutta. Ositteet jaettiin kahtia siten, että 15 suurimman kaupungin terveyskeskuspiirit poimittiin otokseen todennäköisyydellä 1 ja loput 65 systemaattisella PPS-otannalla asukasluvun mukaista suhteellista kiintiöintiä käyttäen. Otokset poimittiin 15 suurimmasta kaupungista suoraan väestömäärän suhteessa. Muista rypäistä otosten poiminta suoritettiin systemaattisella otannalla siten, että väestö oli lajiteltu iän mukaan. 80 vuotta täyttäneiden poimintaväli oli puolet pienempi kuin nuorempien, jotta iäkkäiden määrä tutkimuksessa olisi riittävä. Lopullinen 30 vuotta täyttäneiden osallistujien määrä Terveys 2000 -aineistossa on 8 028. (Heistaro 2005.)

Tilasto-ohjelmistojen rajallisuuden vuoksi otanta-asetelmaa jouduttiin yksinkertaistamaan analyysijä varten kuvan 2.1 kaltaiseksi. Alun perin viidessätoista suurimmassa kaupungissa otanta oli yksivaiheinen ja muissa kaksivaiheinen. Asetelman yksinkertaistaminen koskee ositusta ja ryvästystä yksivaiheisen otannan tapauksessa. Viisitoista suurinta kaupunkia määriteltiin ositteiksi

miljoonapiirien seuraksi, jolloin lopullinen aineisto koostuu 20 ositteesta. Suurimmista kaupungeista poimitut havaintoyksiköt määritettiin rypäiksi samaan tapaan kuin miljoonapiireistä poimitut 65 terveyskeskuspiiriä. (Heistaro 2005.)



Kuva 2.1. Yksinkertaistettu otanta-asetelma.

Otanta-asetelmaa yksinkertaistavien osite- ja ryväsmuuttujien lisäksi aineistoon on lisätty jälkikäteen Tilastokeskuksen muodostamat painokertoimet. Niiden avulla aineiston ikä-, sukupuoli-, alue- ja kielijakaumia muokataan populaation suhteiden mukaisiksi. Väestöpainoa käytetään populaation tunnuslukujen estimoinnissa ja analyysipainoa erilaisten tunnuslukujen ja tilastollisten analyysien estimoinnissa. Lisätietoa aineistoon lasketuista painoista on saatavilla esimerkiksi lähteissä Heistaro (2005) ja Laiho & Nieminen (2004).

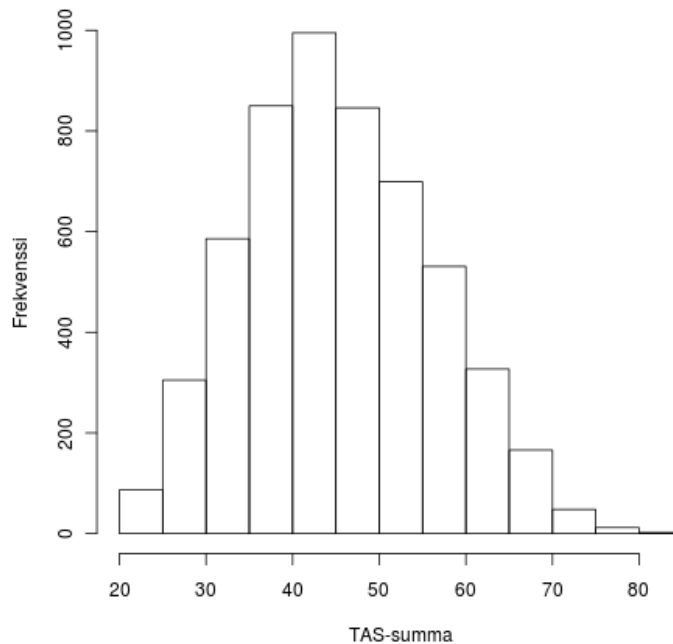
Tämän tutkielman havaintoyksiköiksi valittiin vain ne osallistujat, jotka ovat täyttäneet aleksitymiäkyselyn kokonaisuudessaan ja omalla äidinkielellään. Näin menetellen havaintojen määräksi saadaan 5 454 henkilöä. Havaintojen poistaminen puuttuvien tietojen osalta voi vaikuttaa heikentävästi tulosten yleistettävyyteen. Terveys 2000 -organisaatio suosittelee kaikkien havaintojen käyttämistä ja osajoukon rajaamista tiettyjen tilasto-ohjelmistojen funktioilla. Kuitenkaan erot keskivirheissä eivät ole olleet merkittäviä Terveys 2000 -aineistossa (Heistaro 2005), joten voidaan olettaa, että aineiston rajaaminen ei aiheuta suurta virhettä tuloksiin.

Otantamenetelmä vaikuttaa aineiston analysointiin, sillä saman rypään havaintoyksiköiden välillä voi olla riippuvuutta ja nämä riippuvuussuhteet on huomioitava. Otantamenetelmän ja painotuksen vaikutus vaihtelee eri muuttujien välillä. R-ohjelmiston gam-funktio ei kuitenkaan huomioi otantamenetelmän vaikutusta estimointiin, joten tuloksissa voi ilmetä tästä johtuvaa epätarkkuutta. Otantamenetelmän huomioivien ja perinteisten tunnuslukujen eroja tarkastellaan aleksitymian ja taustamuuttujien esittelyn yhteydessä alaluvuissa 2.2 ja 2.4.1, jotta selviää, kuinka suuri vaikutus menetelmän huomiotta

jättämislä on analysoinnissa.

2.2 Aleksitymia

Aleksitymiaa tarkastellaan kyselyn osioista muodostetun summamuuttujan eli TAS-summan sekä dikotomisen aleksitymia-muuttujan avulla. Aleksityymisiä henkilöitä on aineistossa 555 kpl eli 10,1 %. Summan jakauma näyttää normaalisti jakautuneelta ja sen huippu on noin 45 pisteen kohdalla (ks. kuva 2.2). Summan tunnuslukuja on esitetty taulukossa 2.1 sekä perinteisillä että otantamenetelmän huomioivilla laskukaavoilla laskettuna. Molemmissa tapauksissa keskipistemäärä on noin 46. Erot tunnusluvuissa ovat suhteellisen pienet, joten otantamenetelmän vaikutus analysointituloksiin ei ole kovin suuri.



Kuva 2.2. TAS-summan histogrammi.

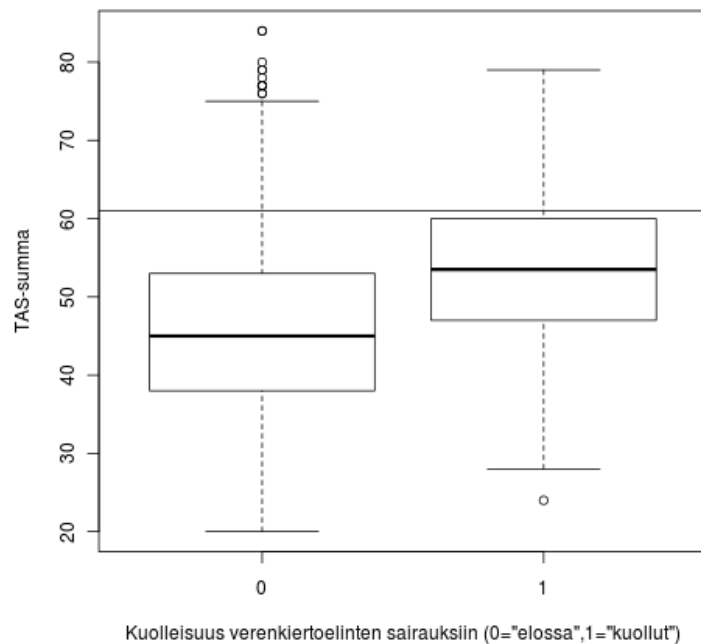
Taulukko 2.1. TAS-summan tunnusluvut.

Otantamenetelmä	Ka	Sd
Ei huomioitu	45,8	10,6
Huomioitu	46,3	10,8

2.3 Kuolleisuus

Kuolleisuus on tutkielman ainoa selitettävä muuttuja. Kuolleisuustietoja oli tutkielman kirjoittamisvaiheessa saatavilla vuoden 2008 loppuun asti. Kuolleisuutta tarkastellaan malleissa sellaisenaan sekä rajattuna kolmeen eri ryhmään kuolinsyiden perusteella. Nämä ryhmät ovat sydän- ja verisuonitautikuolleisuus, verenpaine- tai sydänsairauskuolleisuus ja aivoinfarkti-, verenpaine tai sydänsairauskuolleisuus.

Kuvassa 2.3 on esitetty verenkiertoelinten sairauksiin kuolleiden ja TAS-summan välinen laatikko-janakuvio. Kuvaajan perusteella kuolleiden TAS-summan jakauma on hieman korkeammalla kuin elossa olevien. Aleksityymisten osuus näyttää olevan hieman suurempi kuolleiden kuin elossa olevien joukossa, koska kuolleiden yläkvartiili on lähes aleksityymisyyden rajan kohdalla. Vuoden 2008 loppuun mennessä kuolleita oli yhteensä 501 kpl, joista 112 oli aleksityymisiä (TAS-summa > 60) (ks. taulukko 2.2). Verenpaine- tai iskeemisiin sydänsairauksiin kuolleita on 133 kpl, joista aleksityymisiä 32. Jos lisäksi huomioidaan aivoinfarktidiagnoosit, kuolleita on yhteensä 160, joista aleksityymisiä 40. Aleksityymisistä 20,2 % on kuollut vuoden 2008 loppuun mennessä, kun vastaava osuus ei-aleksityymisillä on vain 7,9 %.



Kuva 2.3. TAS-summan ja verenkiertoelinsairauskuolemien välinen laatikko-janakuvio.

Taulukko 2.2. Aleksityymisyys ja yleinen kuolleisuus.

Aleksityymia	Elossa	Kuollut	Yhteensä
Ei	4 510	389	4 899
On	443	112	555
Yhteensä	4 953	501	5 454

2.4 Taustamuuttajat

2.4.1 Kvantitatiiviset muuttajat

Kuolleisuuteen vaikuttavat useat tekijät, kuten ikä, sukupuoli ja siviilisääty. Nämä taustamuuttajat ovat yhteydessä myös aleksityymisyyteen. Jotta aleksityymian vaikutusta kuolleisuuteen voidaan tutkia, taustamuuttajien vaikutus täytyy huomioida analyyseissä. Tässä alaluvussa käsitellään kvantitatiivisten ja seuraavassa alaluvussa kvalitatiivisten muuttajien tunnuslukuja ja jakaumia. Kvantitatiivisten muuttajien tunnusluvut löytyvät taulukosta 2.3. Svy-alkuisilla muuttujilla viitataan poimintatavan huomioiviin tunnuslukuihin. Erot otantamenetelmän huomioivien ja tavallisten tunnuslukujen välillä eivät ole kovin suuret.

Taulukko 2.3. Taustamuuttajien tunnuslukuja.

Muuttuja	Ka	Sd	Min	1.kvartiili	Md	3. kvartiili	Max
Ikä	51,9	14,5	30,0	40,0	50,0	61,0	97,0
Svy-ikä	51,3	13,7					
Koulutus	11,5	4,1	0,0	8,0	11,0	14,0	33,0
Svy-koulutus	11,5	4,0					
Painoindeksi	26,8	4,6	12,0	25,5	26,2	29,4	54,8
Svy-painoindeksi	26,9	4,6					

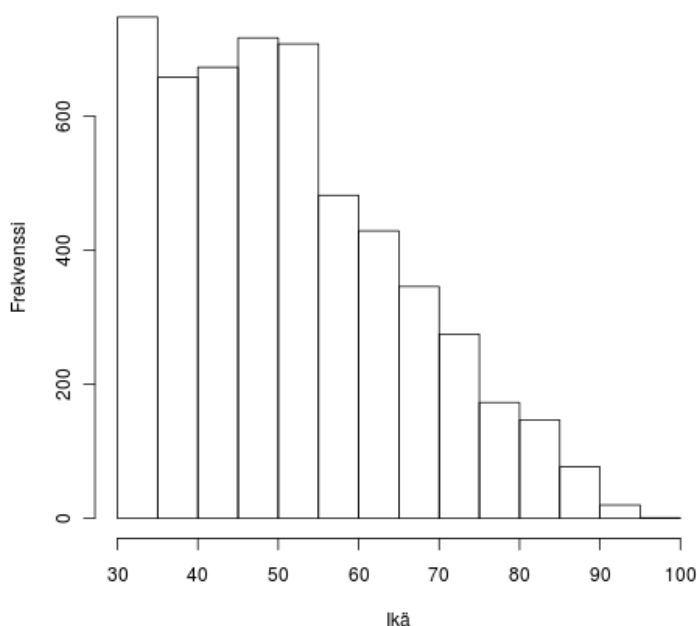
Taulukkoon 2.4 on koottu kvantitatiivisten muuttajien keskiarvot ja -hajonnat erikseen aleksityymisten ja ei-aleksityymisten tapauksissa tarkasteltuna. Koska erot otantatavan huomioivien ja tavallisten tunnuslukujen välillä eivät olleet merkittävän suuria, taulukkoon koottiin vain tavallisten tunnuslukujen arvot. Taulukon mukaan aleksityymisten keski-ikä on kymmenen vuotta suurempi kuin ei-aleksityymisillä. Aleksityymiset opiskelevat keskimäärin kaksi vuotta lyhyemmän ajan kuin ei-aleksityymiset, ja heidän painoindeksinsä on hieman suurempi kuin muilla.

Ikä-muuttuja on henkilön ikä otoksen poimintahetkellä 1.7.2000. Kuvassa 2.4 on esitetty ikäjakauma, joka on oikealle vino. Iän jakauma näyttää pysyttelevän tasaisen korkealla noin 55 vuoden ikään asti ja lähtevän sen jälkeen laskuun. Mediaani-ikä on 50 vuoden kohdalla (ks. taulukko 2.3). Iän tunnusluvuissa ero tavallisen ja otantamenetelmän huomioivan laskutavan välillä on

Taulukko 2.4. Aleksityymisyys ja taustamuuttujien tunnusluvut.

Aleksityymia	Ikä		Koulutus		Painoindeksi	
	Ka	Sd	Ka	Sd	Ka	Sd
Ei	50,9	14,1	11,8	4,0	26,7	4,6
On	60,9	15,4	9,8	3,4	27,9	4,8

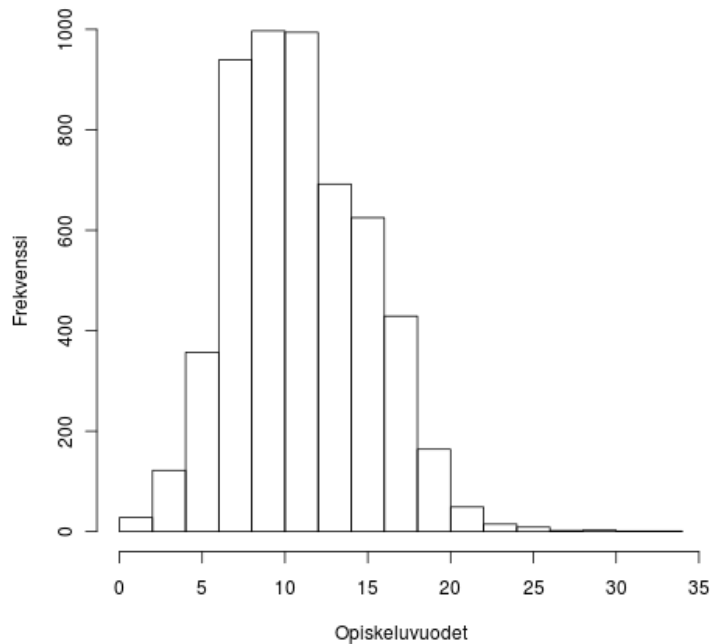
suurin, mutta tässäkin tapauksessa ero keskiarvojen välillä on vain 0,6 ja keskihajontojen välillä 0,8 vuotta.



Kuva 2.4. Iän jakauma.

Koulutusta mitataan opiskeluvuosien lukumäärällä, jonka frekvenssihistogrammi on kuvassa 2.5. Jakauman huippu sijoittuu melko leveälle aikavälille kuudesta kahteentoista vuoteen. Suosituimmat opintoajat ovat siis ala-asteen, yläasteen tai toisen asteen koulutusten pituisia. Taulukosta 2.3 nähdään, että keskimääräinen opiskeluaika on 11–12 vuotta. Alin neljännes opiskelee korkeintaan 8 vuotta, kun taas ylin neljännes opiskelee vähintään 14 ja enintään 33 vuotta.

Painoindeksi kuvaa henkilön painon ja pituuden suhdetta ja se lasketaan jakamalla paino (kg) pituuden neliöllä (m^2). Henkilö on normaalipainoinen, jos painoindeksin arvo on yli 18,5 ja alle 25. Välillä 25–30 olevat painoindeksin arvot viittaavat lievään lihavuuteen ja lukua 40 suuremmat arvot sairaalloyeen lihavuuteen. (Mustajoki 2011.) Taulukon 2.3 tunnusluvuista nähdään, että pai-



Kuva 2.5. Opiskeluvuosien lukumäärän jakauma.

noindeksi on suurimmalla osalla yli normaalin viitearvojen. Yhteensä 75 %:lla havainnoista painoindeksi viittaa jonkin asteiseen lihavuuteen. Keskimmaisilla 50 %:lla painoindeksi on lievän lihavuuden alueella välillä 25,5–29,4.

2.4.2 Kvalitatiiviset muuttujat

Miehiä on aineistossa 2 480 (45,5 %) ja naisia 2 974 (54,5 %). Kvalitatiivisten muuttujien %-osuudet aleksityymisten, ei-aleksityymisten ja kaikkien havaintojen joukossa on esitetty taulukossa 2.5. Sivilisäyty on jaettu kahteen luokkaan siten, että avio- tai avoliitossa olevat on koodattu nolllaksi ja muut ykköseksi. Havaintoyksiköistä 71,0 % on parisuhteessa. Aleksityymisistä parisuhteessa on noin kymmenen %-yksikköä pienempi osuus kuin ei-aleksityymisistä. Vapaa-ajan liikuntaa harrastaa 73 % vastaajista, mutta aleksityymisistä vain 57,8 %. Vastaajista 21,7 % tupakoi säännöllisesti. Aleksityymisten joukossa säännöllisesti tupakoivien osuus on 23,6 % eli hieman suurempi kuin kaikkien vastanneiden keskuudessa. Vastaajista 4,1 %:lla on diagnosoitu alkoholihäiriö. Häiriöiksi lasketaan alkoholin väärinkäyttö ja alkoholiriippuvuus. Alkoholihäiriödiagnoosi on 5,6 %:lla aleksityymisistä.

Masennus ja ahdistuneisuus näyttää olevan selvästi yleisempää aleksityymisten joukossa. Masennusta kuvataan kaksiluokkaisella depressio-muuttujalla, ja sen perusteella 6,5 %:lla vastaajista on masennusdiagnoosi. Aleksityymisistä 14,0 %:lla on masennus, kun ei-aleksityymisten vastaava osuus on vain 5,8 %. Ahdistuneita on aineistossa 4,1 % ja osuus on 10,0 %, kun tarkastellaan ai-

noastaan aleksityymiä. Metabolista oireyhtymää sairastaa 24,6 % havaintoyksiköistä ja aleksityymisten joukossa osuus on vieläkin suurempi (37,5 %).

Taulukko 2.5. Kvalitatiivisten muuttujien %-osuudet aleksityymisten ja ei-aleksityymisten joukoissa.

Aleksityymia	Miehiä (%)	Parisuhde (%)	Liikunta (%)	Tupakointi (%)
Ei	44,5	72,1	74,7	21,5
On	54,1	62,0	57,8	23,6
Kaikki	45,5	71,0	73,0	21,7
	Alkoholi (%)	Depressio (%)	Ahdistus (%)	Metabolinen (%)
Ei	4,0	5,8	3,5	23,2
On	5,6	14,0	10,0	37,5
Kaikki	4,1	6,5	4,1	24,6

3 Tutkimusmenetelmät

3.1 Yleistetyt lineaariset mallit (GLM)

Tässä aluvuussa on käytetty lähteinä pääasiassa Alan Agrestin kirjaa ”An Introduction to Categorical Data Analysis” (2007) ja Jarkko Isotalon luentomonistetta Yleistetyt lineaariset mallit I -kurssilta (2009). Yleistetyissä lineaarisissa malleissa selittävien muuttujien vaikutusta vasteeseen mallinnetaan odotusarvon ja linkkifunktion avulla.

Olkoon mallissa m selittävää muuttujaa X_1, X_2, \dots, X_m ja n havaintoa. Satunnaismuuttujia vastaavat havaitut arvot ovat $x_{1i}, x_{2i}, \dots, x_{mi}$, missä i viittaa i :n havainnon saamiin arvoihin. Olkoon lisäksi Y_i selitettävä satunnaismuuttuja ja merkitään sen odotusarvoa $E(Y_i) \equiv \mu_i$. Merkitään linkkifunktiota g :llä. Tällöin malli on muotoa

$$(3.1) \quad g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi},$$

missä $i = 1, 2, \dots, n$, β_0 on vakiotermi ja $\beta_1, \beta_2, \dots, \beta_m$ ovat kerrointermejä.

Selitettävän muuttujan Y oletetaan noudattavan jotain eksponentiaaliseen perheeseen kuuluvaa jakaumaa. Tällaisia jakaumia ovat esimerkiksi normaali-, binomi-, Poisson- ja gammajakaumat. Lisäksi oletetaan, että linkkifunktio on monotoninen ja derivoituva. Linkkifunktio valitaan selitettävän muuttujan jakauman perusteella. Mahdollisia linkkifunktioita ovat muun muassa identiteetti-, log- ja logit-linkit.

Identiteettilinkki on yksinkertainen linkkifunktio, ja se on muotoa $g(\mu) = \mu$. Jos jakaumaoletuksena on normaalijakauma ja valitaan identiteettilinkki, saadaan erikoistapauksena tavallinen lineaarinen malli. Jos selitettävä muuttuja saa positiivisia kokonaislukuarvoja, sen mallintamiseen voidaan käyttää Poisson-jakaumaa. Tällöin linkkifunktioksi sopii log-linkki ja mallia kutsutaan log-lineaariseksi regressiomalliksi. Jos selitettävän muuttujan odotusarvo voi saada arvoja vain väliltä $[0, 1]$, sopii linkkifunktioksi logit-linkki ja jakaumaksi binomijakauma. Tällaista mallia kutsutaan usein logistiseksi regressiomalliksi. Koska tutkielman analyyseissä käytetään binomijakaumaoletusta ja logit-linkkiä, niitä tarkastellaan seuraavaksi tarkemmin.

Jos selitettävä muuttuja on dikotominen eli sillä on kaksi toisensa poissulkevaa tulostulovaihtoehtoa, voidaan käyttää logistista regressiomallia. Muuttuja koodataan nolllaksi ja ykköseksi, missä ykkönen kuvaa onnistumista. Tällöin odotusarvon arvoalue on välillä $[0, 1]$ ja mallinnuksen kohteena on onnistumisen todennäköisyys

$$\pi_i = P(Y_i = 1 | X_{1i} = x_{1i}, X_{2i} = x_{2i}, \dots, X_{mi} = x_{mi}).$$

Onnistumistodennäköisyydestä voidaan muodostaa vedonlyöntisuhde:

$$\gamma_i = \frac{\pi_i}{1 - \pi_i}.$$

Kun vedonlyöntisuhteesta otetaan luonnollinen logaritmi, saadaan onnistumistodennäköisyyden logit-muunnos. Logistisen regression tapauksessa malli muodostetaan onnistumistodennäköisyyden logit-muunnokselle. Tällöin linkkifunktio g on siis logit-linkki:

$$\begin{aligned} g(\mu_i) &= \log(\gamma_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \text{logit}(\pi_i) \\ &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi}. \end{aligned}$$

Yleistettyjen lineaaristen mallien estimoinnissa käytetään useimmiten suurimman uskottavuuden menetelmää. Tilastollinen päättely tapahtuu parametrien estimaattien ja p-arvojen tarkastelulla sekä kuvaajien avulla. Mallin hyvyttä voidaan arvioida esimerkiksi Akaiken informaatiokriteerin (*AIC*), residuaalivertailujen ja hyvyysindeksien avulla. Lisätietoa estimoinnista ja päätelystä on esitetty monissa lähteissä, kuten esimerkiksi Simon N. Woodin teoksessa *Generalized Additive Models: An Introduction with R* (2006).

3.2 Yleistetyt additiiviset mallit (GAM)

Tässä ja seuraavissa alaluvuissa esitetty teoria perustuu pääasiassa David Ruppertin, M. P. Wandin ja R. J. Carrollin kirjaan ”Semiparametric Regression” (2003) ja Simon N. Woodin kirjaan ”Generalized Additive Models: An Introduction With R” (2006). Yleistetyt additiiviset mallit ovat yleistettyjen lineaaristen mallien yleistys, joiden etuna on epälineaaristen yhteyksien joustavampi mallintaminen. Yleistetyissä lineaarisissa malleissa epälineaarisia yhteyksiä voidaan mallintaa esimerkiksi sopivilla muuttujamuunnoksilla tai useamman asteen polynomeilla. Jos mallinnettava riippuvuus on monimutkainen ja voimakkaasti epälineaarinen, sopivan mallirakenteen valinta voi muodostua hankalaksi. Yleistetyissä additiivisissa malleissa ongelma korjaantuu siten, että tarvittavat muunnokset estimoidaan automaattisesti sopivien tasoitusfunktioiden avulla. (Guisan, Edwards & Hastie 2002.)

Yleistettyjen additiivisten mallien rakenne on samankaltainen kuin yleistettyjen lineaaristen mallien, mutta lisäksi selittäjiä voidaan mallintaa epäparametrisesti tasoittavien funktioiden (*smooth functions, f*) kautta. Esimerkiksi mallin (3.1) β -parametrit voidaan korvata tasoittavilla funktioilla, jolloin saadaan seuraava malli:

$$(3.2) \quad g(\mu_i) = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_l(x_{mi}), \quad \text{missä } i = 1, 2, \dots, n.$$

Oletukset sekä linkkifunktion ja vasteen jakauman valinnat ovat samanlaiset kuin yleistetyissä lineaarisissa malleissa. Binomijakaumaoletuksella ja logit-linkin valinnalla malli (3.2) on muotoa

$$(3.3) \quad \text{logit}(\pi_i) = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_m(x_{mi}), \quad \text{missä } i = 1, 2, \dots, n.$$

Tasoittavia funktioita voidaan mallintaa monilla eri tasoitusmenetelmillä, kuten esimerkiksi splineillä, Kernel-tasoituksella tai muilla paikallisilla polynomisovitteilla (*local polynomial fitting*). (Eubank 1999; Hastie & Tibshirani 1990.) Tämän tutkielman analyyseissä tasoitusmenetelminä on käytetty sakoitettuja ”thin plate” -regressiosplinejä ja interaktiitermien kohdalla tensoritulotasoitteita (*tensor product smooths*), joita esitellään tarkemmin seuraavassa alaluvussa.

Mallin valinta perustuu samoihin menetelmiin kuin yleistettyjen lineaaristen mallien tapauksessa. Kahden mallin vertailu onnistuu esimerkiksi testaamalla χ^2 -testillä devianssien erotuksen merkitsevyyttä tai vertaamalla Akaiken informaatiokriteereitä. Tässä tutkielmassa mallien vertailussa käytetään devianssien erotuksen merkitsevyyden testausta. Tilastollinen päättely perustuu käytännössä kuvaajan tulkintaan sekä termien merkitsevyyksien ja mallin hyvyyden tarkasteluun. Tasoittavalle funktiolle määritetään 95 %:n luottamusväli, joka esitetään mallin soviteen kuvaajissa. Tilastollisen päättelyn teoreettista perustelua ja muita mallinvalintakriteereitä on esitetty kattavasti kirjallisuudessa (esim. Eubank 1999; Ruppert, Wand & Carroll 2003).

GAM-malleissa osa selittävästä muuttujista voi olla luonteeltaan lineaarisia. Silloin niiden mallintaminen tasoittavien funktioiden avulla ei ole järkevää. Siksi mallissa sallitaan myös lineaariset termit. Olkoon mallissa l kpl lineaarisesti ja $(m - l)$ kpl epälineaarisesti malliin tulevia selittäjiä. Nyt malli (3.3) saadaan muotoon:

$$(3.4) \quad \begin{aligned} \text{logit}(\pi_i) = & \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_l x_{li} \\ & + f_{l+1}(x_{(l+1)i}) + \dots + f_m(x_{mi}), \quad \text{missä } i = 1, 2, \dots, n. \end{aligned}$$

Selittäjiä voidaan siirtää tasoitettavien termien puolelta lineaariselle, jos mallin antamat tulokset viittaavat sen tarpeellisuuteen. Jokaiselle tasoitetulle termille lasketaan efektiiviset vapausasteet (*edf*), joiden avulla päättely muuttujan lineaarisuudesta voidaan tehdä. Efektiiviset vapausasteet perustuvat solmukohtien (*knots*) lukumäärään. Solmukohtia käsitellään tarkemmin seuraavassa alaluvussa 3.3.1. Vapausasteita voi olla enimmillään yksi vähemmän kuin solmukohtia. Lineaarisen termin sovite kulkee kahden pisteen eli solmukohdan kautta, jolloin vapausasteita on yksi. Jos efektiiviset vapausasteet ovat siis lähellä ykköstä, muuttujan sovite on lineaarinen, eikä se tarvitse tasoitusta. Epälineaarilla termeillä vapausasteita on enemmän, ja niiden määrä riippuu siitä, kuinka kaareva tasoittava funktio on. Vapausasteet saadaan ratkaistua matriisin jäljen avulla. Olkoon \mathbf{P}_i i :nmen tasoitteen parametrit palauttava matriisi

ja \mathbf{X} mallimatriisi. Nyt i :n termin vapausasteet saadaan jäljestä $tr(\mathbf{X}\mathbf{P}_i)$. Kun termien vapausasteet lasketaan yhteen, saadaan koko mallin vapausasteet. (Wood 2012.)

Joskus selittävien muuttujien yhteisvaikutus vasteeseen on suurempi kuin yksittäisten vaikutusten summa. Esimerkiksi sairauden riskitekijöiden kasautuminen lisää sairastumisen todennäköisyyttä enemmän kuin yksittäisten riskitekijöiden vaikutusten verran. Tätä ilmiötä kutsutaan interaktioksi. Interaktio on siis voimassa silloin, kun additiivisuus ei toteudu. Interaktio voidaan kuitenkin huomioida GAM:eissa lisäämällä malliin yhdysvaikutustermejä $f_{ij}(x_i, x_j)$. Yhdysvaikutustermit voidaan lisätä esimerkiksi malliin (3.4) kaikkien tasoittavien funktioiden kautta mallinnettavien selittävien muuttujien välille. Näin saadaan malli (3.5):

$$(3.5) \quad \begin{aligned} \text{logit}(\pi_i) = & \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_l x_{li} \\ & + f_{l+1}(x_{(l+1)i}) + \dots + f_m(x_{mi}) \\ & + f_{(l+1)(l+2)}(x_{(l+1)i}, x_{(l+2)i}) + \dots \\ & + f_{(m-1)m}(x_{(m-1)i}, x_{mi}), \end{aligned}$$

missä $i = 1, 2, \dots, n$.

3.3 Tasoitusmenetelmät

3.3.1 Johdatus splineihin

Splinillä tarkoitetaan käyrää, joka on muodostettu yhdistämällä toisiinsa useita pienemmällä osaväleillä määriteltyjä lineaarisia funktioita. Osavälit yhdistyvät toisiinsa solmukohtien (κ) avulla. Splini muodostetaan siis paloittain määritellyistä lineaarisista funktioista, jotka liittyvät toisiinsa solmukohdissa. Splinille on määritettävä kanta ja sitä vastaava funktioavaruus. Kantafunktiot voidaan määrittää selittävien muuttujien arvojen sopivalla muunnoksella, kuten esimerkiksi logaritminuunnoksilla tai polynomiregressiolla (Nummi 2008). Seuraavassa esityksessä kantafunktiona käytetään tyypistettyä potenssikantaa.

Tarkastellaan yksinkertaista epäparametrasta mallia

$$(3.6) \quad y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \dots, n,$$

missä y_i on selitettävä ja x_i selittävä muuttuja, f on tasoittava funktio ja ϵ_i on jäännöstermi. Jäännöstermi kuvaa satunnaisvaihtelua, jota ei mallin avulla pystytä selittämään. Oletuksena on, että jäännöstermit ovat riippumattomia, normaalijakaumaa noudattavia satunnaismuuttujia, joiden odotusarvo on nolla ja varianssi σ^2 on vakio.

Olkoot mallilla (3.6) solmut $\kappa_1, \dots, \kappa_K$, jotka on määritelty havaintovälillä $[1, n]$ siten, että $1 < \kappa_1 < \dots < \kappa_K < n$. Nyt ensimmäisen asteen tyypistetyt kantafunktiot ovat muotoa $1, x, (x - \kappa_1)_+, \dots, (x - \kappa_K)_+$. Funktiot

$(x - \kappa_1)_+, \dots, (x - \kappa_K)_+$ on määritelty vain positiivisilla arvoalueillaan eli kun $x > \kappa_k$, missä $k = 1, 2, \dots, K$. Muulloin funktiot saavat arvon nolla.

Kantafunktioiden lineaarikombinaationa saadaan tasoittavalle funktiolle seuraavanlainen kaava:

$$(3.7) \quad f(x) = \beta_0 + \beta_1 x + b_1(x - \kappa_1)_+ + \dots + b_K(x - \kappa_K)_+,$$

kun $\beta_0, \beta_1, b_1, \dots, b_K$ ovat kantafunktioiden muunnoskertoimia. Sijoittamalla funktio (3.7) kaavaan (3.6), saadaan lineaarinen malli

$$y_i = \beta_0 + \beta_1 x_i + b_1(x_i - \kappa_1)_+ + \dots + b_K(x_i - \kappa_K)_+ + \epsilon_i,$$

missä $i = 1, 2, \dots, n$. Mallin sovite lasketaan pienimmän neliösumman avulla, mutta lisäksi tyypistettyjen kantafunktioiden kertoimia b_1, \dots, b_K voidaan rajoittaa, jotta vältetään ylisovittamiselta.

R-ohjelmiston mgcv-kirjastossa on saatavilla useita eri tasoitusmenetelmiä (Wood 2006), mutta koska tämän tutkielman analyyseissä käytetään regressiosplinejä, tarkastellaan niitä seuraavaksi tarkemmin. Yleisiä regressiosplinejä ovat esimerkiksi kuutiolliset ja thin plate -regressiosplinit. Kuutiolliset regressiosplinit voidaan muodostaa käyttämällä kolmannen asteen tyypistettyä kantaa $1, x, x^2, x^3, (x - \kappa_1)_+^3, \dots, (x - \kappa_K)_+^3$, missä $\kappa_1, \dots, \kappa_K$ ovat funktion solmukohdat. Nyt kuutiollinen splini voidaan kirjoittaa muodossa

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + b_1(x - \kappa_1)_+^3 + \dots + b_K(x - \kappa_K)_+^3,$$

missä $\beta_0, \beta_1, \beta_2, \beta_3, b_1, \dots, b_K$ ovat kantafunktioiden kerroinparametrit. Kuutiollisten regressiosplinin estimointiin voidaan käyttää samoja menetelmiä kuin lineaarisissa malleissa (esim. Hastie & Tibshirani 1990).

Solmukohtien määrän ja sijainnin valinnalla on tärkeä merkitys monien regressiosplinin sovituksen onnistumisessa. Jos solmuja valitaan liikaa, aineiston satunnaisvaihtelu saa liian suuren huomion, ja sovitteesta tulee liian tarkka. Jos solmukohtia on liian vähän, sovite jättää tärkeätkin yksityiskohdat huomiotta, ja malli on aliestimoitu.

Solmukohtien määrä ja sijainti voidaan valita monin eri tavoin, mutta laskennallisten syiden vuoksi määrä on syytä pitää suhteellisen pienenä. Hyvä nyrkkisääntö on, että kaikkien solmujen väliin jäisi ainakin 5 havaintoa, ja kuitenkin suurissakin aineistoissa olisi korkeintaan 20–40 solmua. Solmujen määrän voi valita esimerkiksi tasavälein tai kuvaajan perusteella, mutta kirjallisuudessa on esitetty valintaan myös erilaisia algoritmeja (esim. Nummi 2008; Ruppert et al. 2003).

Koska solmujen valinnalla on suuri vaikutus regressiosplinitasoitteen onnistumiseen, on syytä esitellä myös vaihtoehtoinen tapa tasoittaa funktiota. Tasoittavien splinin avulla vältetään solmujen määrän ja sijainnin määrittämiseltä, koska kaikki havainnot toimivat solmukohtina. Jokaisen havainnon

huomioiva funktio on kuitenkin hyvin rosoinen, joten sitä voidaan tasoittaa sakkotermin avulla. Nyt määritettäväksi jää vain sakkotermin sisältävä tasoitusparametri λ , jonka avulla tasaisuutta voidaan kontrolloida.

Kuten solmukohtien määrän ja sijainnin valinnalla, myös tasoitusparametrin valinnalla on suuri vaikutus sovituksen onnistumisen kannalta. Pienillä λ :n arvoilla käyrä menee tarkasti havaintopisteiden mukaisesti. Rosoisuutta saadaan siis tasoitettua valitsemalla sopivan suuri λ . On kuitenkin varottava valitsemasta liian suurta arvoa, sillä käyrä lähenee suoraa, kun λ lähenee ääretöntä. Tasoitusparametri voidaan solmukohtien tapaan valita silmämääräisesti kuvaajien perusteella tai käyttäen erilaisia automaattisia menetelmiä. Automaattisia valintamenetelmiä ovat esimerkiksi ristiinvalidointi (*cross-validation*, *CV*), yleistetty ristiinvalidointi (*generalized cross-validation*, *GCV*), *Mallowsin* C_p -kriteeri tai Akaiken ja Bayesin informaatiokriteerit.

3.3.2 Thin plate -regressiosplinit

Edellisessä luvussa todettiin solmukohtien valinnalla olevan suuri merkitys regressiospliniin onnistuneessa estimoinnissa. Thin plate -regressiospliniin avulla solmukohtien valinnalta kuitenkin vältytään. Lisäksi ne soveltuvat usean selittäjän tasoitteiden estimointiin eli niiden avulla malliin voi lisätä interaktiotermiä.

Oletetaan, että estimoitavana on kaavan (3.6) mukaisen mallin tasoittava funktio f . Thin plate -splinit estimoivat funktion etsimällä sellaisen estimaatin \hat{g} , joka minimoi lausekkeen

$$\| \mathbf{y} - \mathbf{g} \|^2 + \lambda J_{md}(g),$$

missä \mathbf{y} on selitettävien arvojen y_i vektori ja $\mathbf{g} = [g(x_1), g(x_2), \dots, g(x_n)]^T$. $J_{md}(g)$ on sakkofunktio, joka mittaa funktion g rosoisuutta, ja λ on tasoitusparametri. Minimointiteoriaa on esitetty tarkemmin useissa teoksissa (esim. Wood 2006). Thin plate -regressiosplinit perustuvat ajatukseen, että thin plate -spliniin rosoisuuskomponentteja typistetään, kun muut komponentit pysyvät ennallaan.

Vaikka thin plate -splineillä on monia etuja muihin regressiosplineihin verrattuna, ne eivät sovellu kaikkiin tilanteisiin. Etenkin suurilla aineistoilla ne eivät ole laskennallisesti tehokkaita, koska estimoitavia parametreja on yhtä paljon kuin havaintoja. Lisäksi ne eivät sovellu käytettäväksi interaktiotarkasteluissa, jos muuttujat on mitattu eri asteikoilla.

3.3.3 Tensoritulotasoihte

Jos halutaan mallintaa interaktiota sellaisten muuttujien välillä, jotka eivät ole luonnollisesti samalla asteikolla, voidaan käyttää tensoritulotasoitetta. Tensoritulon kantana voidaan käyttää mitä tahansa kantojen yhdistelmää. Rohjelmiston *mgcv*-kirjastossa kantoina voidaan käyttää kaikkia tarjolla olevia splinejä, mutta tarvittaessa kannat voi määrittää myös itse. Tämän tutkielman

interaktioiden kantana käytetään vain thin plate -regressiosplinejä. Koska tensoritulot ovat solmuperusteisia, thin plate -regressiospliniin solmottomuuteen perustuva tehokkuus ei kuitenkaan tuota tensoritulotasoituksessa etua muihin kantoihin verrattuna.

Oletetaan, että halutaan analysoida malli, jossa on kahden selittävän muuttujan x ja z lisäksi niiden välinen interaktiotermin. Olkoon f_x ja f_z muuttujien tasoittavat funktiot ja a_i ja b_j niiden tunnetut kantafunktiot. Nyt funktiot voidaan kirjoittaa muodossa

$$f_x(x) = \sum_{i=1}^I \alpha_i a_i(x) \quad \text{ja} \quad f_z(z) = \sum_{j=1}^J \beta_j b_j(z),$$

missä α_i ja β_j ovat kerroinparametrit.

Muodostetaan nyt funktion f_x avulla interaktioiden tasoittava funktio f_{xz} . Tämä onnistuu sallimalla muunnoskerroinparametrin α_i arvojen vaihdella tasaisesti muuttujan z arvojen kanssa. Käyttämällä muuttujan z tasoittavan funktion esityksessä käytettyä kantaa, voidaan määrittellä

$$\alpha_i(z) = \sum_{j=1}^J \beta_{ij} b_j(z),$$

jolloin interaktion tasoittava funktio voidaan kirjoittaa muodossa

$$f_{xz}(x, z) = \sum_{i=1}^I \sum_{j=1}^J \beta_{ij} b_j(z) a_i(x).$$

Tensoritulotasoitteen määrittävän mallimatriisin \mathbf{X} ja muuttujien reunatasoitteet määrittävien mallimatriisien \mathbf{X}_x ja \mathbf{X}_z yhteys voidaan esittää *Kroneckerin tulon* (\otimes) avulla. Nyt i :nnes rivi mallimatriisista \mathbf{X} on muotoa

$$\mathbf{X}_i = \mathbf{X}_{xi} \otimes \mathbf{X}_{zi}.$$

Kroneckerin tulo on yleistys vektoreiden ulkotulosta matriiseihin. Jos X_x on $(n \times m)$ -matriisi ja X_z $(p \times q)$ -matriisi, niiden välinen Kroneckerin tulo on $(np \times qm)$ -matriisi:

$$\mathbf{X}_x \otimes \mathbf{X}_z = \begin{bmatrix} X_{x(11)} \mathbf{X}_z & X_{x(12)} \mathbf{X}_z & \dots & X_{x(1m)} \mathbf{X}_z \\ X_{x(21)} \mathbf{X}_z & X_{x(22)} \mathbf{X}_z & \dots & X_{x(2m)} \mathbf{X}_z \\ \vdots & \vdots & \ddots & \vdots \\ X_{x(n1)} \mathbf{X}_z & X_{x(n2)} \mathbf{X}_z & \dots & X_{x(nm)} \mathbf{X}_z \end{bmatrix},$$

eli tarkemmin

$$\mathbf{X}_x \otimes \mathbf{X}_z =$$

$$\begin{bmatrix} X_{x(11)}X_{z(11)} & \cdots & X_{x(11)}X_{z(1q)} & \cdots & \cdots & X_{x(1m)}X_{z(11)} & \cdots & X_{x(1m)}X_{z(1q)} \\ \vdots & \ddots & \vdots & \cdots & \cdots & \vdots & \ddots & \vdots \\ X_{x(11)}X_{z(p1)} & \cdots & X_{x(11)}X_{z(pq)} & \cdots & \cdots & X_{x(1m)}X_{z(p1)} & \cdots & X_{x(1m)}X_{z(pq)} \\ \vdots & & \vdots & \ddots & & \vdots & & \vdots \\ \vdots & & \vdots & & \ddots & \vdots & & \vdots \\ X_{x(n1)}X_{z(11)} & \cdots & X_{x(n1)}X_{z(1q)} & \cdots & \cdots & X_{x(nm)}X_{z(11)} & \cdots & X_{x(nm)}X_{z(1q)} \\ \vdots & \ddots & \vdots & \cdots & \cdots & \vdots & \ddots & \vdots \\ X_{x(n1)}X_{z(p1)} & \cdots & X_{x(n1)}X_{z(pq)} & \cdots & \cdots & X_{x(nm)}X_{z(p1)} & \cdots & X_{x(nm)}X_{z(pq)} \end{bmatrix}.$$

4 GAM:in soveltaminen aineistoon

4.1 Mallin sovittaminen

Tutkielmassa mallinnetaan yleistä kuolleisuutta sekä kolmea kuolinsyiden perusteella rajattua pienempää kuolleiden joukkoa aleksitymiäkyselyn summa-
muuttujalla ja useilla taustamuuttujilla. Taustamuuttujista suurin osa lisätään
malliin lineaarisesti, koska ne ovat dikotomisia luokitteluasteikollisia muuttu-
jia. Jatkuvat muuttujat lisätään malliin tasoittavien funktioiden kautta. Täl-
löin aineistoon sovitettava malli on seuraavan kaltainen:

$$(4.1) \quad \begin{aligned} \text{logit}(\pi) = & \beta_0 + \beta_1 \text{sukupuoli} + \beta_2 \text{siviilisäätö} + \beta_3 \text{depressio} \\ & + \beta_4 \text{ahdistus} + \beta_5 \text{alkoholi} + \beta_6 \text{tupakka} \\ & + \beta_7 \text{liikunta} + \beta_8 \text{metabolinen} \\ & + f_1(\text{TAS-summa}) + f_2(\text{ikä}) + f_3(\text{koulutus}) + f_4(\text{bmi}). \end{aligned}$$

Mallia voidaan parannella esimerkiksi poistamalla mahdollisia ei-merkitseviä
selittäjiä tai siirtämällä tarvittaessa tasoittavien funktioiden kautta mallinnet-
tavia muuttujia lineaarisiksi selittäjiksi. Lineaarisuus voidaan havaita efektiiv-
isten vapausasteiden arvon perusteella tai kuvaajasta. Jos muuttujan vaikutus
vasteeseen on lineaarinen, estimoidun tasoittavan funktion kuvaaja on suora.

Aleksitymian yhteys kuolleisuuteen voi olla todellisuudessa välillistä eli vai-
kutuksen voi ilmetä muiden muuttujien kautta. Kun esimerkiksi aleksitymial-
la on yhteys alkoholin runsaaseen käyttöön ja alkoholi on yleinen kuolinsyy,
on mahdollista, että aleksitymia vaikuttaa kuolleisuuteen myös alkoholin kaut-
ta. GAM-malleilla tätä epäsuoraa yhteyttä voidaan selvittää yhdysvaikutuster-
mien avulla. Tässä tutkielmassa keskitytään tasoitettujen termien välisten inte-
raktioiden tutkimiseen. Yhdysvaikutustermejä voidaan siis lisätä TAS-summan
ja iän, koulutuksen sekä painoindeksin välille. Kun malliin (4.1) lisätään kaikki
kolme yhdysvaikutustermiä, interaktiomalli on muotoa

$$(4.2) \quad \begin{aligned} \text{logit}(\pi) = & \beta_0 + \beta_1 \text{sukupuoli} + \beta_2 \text{siviilisäätö} + \beta_3 \text{depressio} \\ & + \beta_4 \text{ahdistus} + \beta_5 \text{alkoholi} + \beta_6 \text{tupakka} \\ & + \beta_7 \text{liikunta} + \beta_8 \text{metabolinen} \\ & + f_1(\text{TAS-summa}) + f_2(\text{ikä}) + f_3(\text{koulutus}) + f_4(\text{bmi}) \\ & + f_5(\text{TAS-summa,ikä}) + f_6(\text{TAS-summa,koulutus}) \\ & + f_7(\text{TAS-summa,bmi}). \end{aligned}$$

Kuten varsinaisten yleistettyjen additiivisten mallienkin kohdalla, interaktiomallia voidaan parannella poistamalla siitä ei-merkitseviä selittäjiä tai yhdysvaikutustermejä sekä siirtämällä selittäjiä tarvittaessa lineaarisen mallintamisen puolelle.

4.2 Tulokset

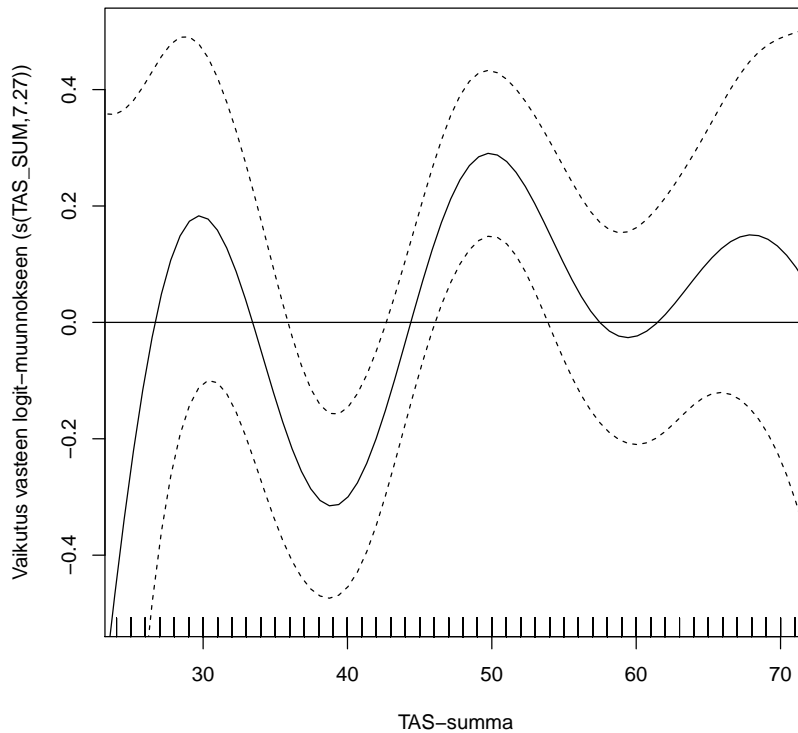
Kaikki mallit on muodostettu siten, että alkuasetelma on kaavan (4.1) mukainen. Tasoittavien funktioiden kantana on käytetty thin plate -regressiosplinejä. Alkuperäistä mallia on muokattu tarvittaessa poistamalla tasoitus sellaisten muuttujien kohdalta, joiden vaikutus vasteeseen on lineaarinen. Ei-merkitseviä termejä ei sen sijaan ole poistettu malleista, koska tavoitteena on selvittää aleksitymian yhteyttä kuolleisuuteen, kun taustamuuttujien vaikutus on huomioitu. Samasta syystä pääpaino tulosten esityksessä on aleksitymiaa kuvaavan TAS-summamuuttujan ja kuolleisuuden välisen yhteyden tutkimisessa.

Tarkastellaan aluksi mallia, jossa selitettävänä muuttujana on yleinen kuolleisuus. Alkuperäisessä mallissa koulutukselle estimoidut efektiiviset vapausasteet olivat lähellä yhtä ($edf=1,005$), joten muuttujan mallintaminen tasoittavilla funktioilla ei ollut mielekäästä. Lopullisessa mallissa koulutus on siis lisätty lineaariseksi selittäjäksi ja tasoittavien funktioiden avulla malliin tulevat vain TAS-summa, ikä ja painoindeksi. Mallin estimaatit p-arvoineen on esitetty liitteen B taulukossa 1. Koulutusta ($p=0,79$) ja alkoholin kulutusta ($p=0,09$) lukuun ottamatta kaikki termit ovat tilastollisesti merkitseviä 5 %:n riskitasolla. TAS-summalla ja painoindeksillä on runsaat seitsemän efektiivistä vapausastetta (7,27 ja 7,25) ja iällä 2,55. Mallin avulla devianssista saadaan selitettyä 29,4 %.

TAS-summan vaikutus kuolleisuuden logit-muunnokseen on esitetty kuvassa 4.1. Kuvassa näkyy tasoittavan funktion estimaatti yhtenäisenä käyränä. Katkoviivakäyrät estimaatin ympärillä kuvaavat funktion 95 %:n luottamusväliä. Havaintojen määrä vaikuttaa luottamusvälin leveyteen, ja siksi välit levenevät havaintovälin päätepisteiden läheisyydessä, joissa havaintoja on vähemmän. Tästä syystä kuvaajia on rajattu siten, että aivan pienimmät ja suurimmat summan arvot eivät näy kuvassa. TAS-summamuuttujan arvot ovat x-akselilla ja y-akselilla on summamuuttujan vaikutus kuolleisuuden logit-muunnokseen, kun muiden muuttujien vaikutus on vakioitu. Kuvaajan alalaidassa x-akselille piirretyt viivat ovat TAS-summan havaitut arvot. Viivan paksuus viittaa havaintojen määrään siten, että paksumman viivan kohdalla on enemmän havaintoja kuin ohuen viivan kohdalla.

Kuvaa tulkitaan y-akselin nollakohdan avulla. Jos käyrä luottamusvälineen jää nollakohdan ala- tai yläpuolelle, on muuttujan vaikutus vasteeseen tilastollisesti merkitsevä. Muuttujan vaikutus on positiivinen, jos käyrä on nollan yläpuolella ja vastaavasti negatiivinen nollan alapuolella. Käyrä aaltoilee nollan molemmin puolin, ja sillä on kolme huippua suurin piirtein TAS-summan arvojen 30, 50 ja 70 kohdilla. Kuvaajan perusteella TAS-summalla näyttää olevan

kahdella pienellä alueella merkitsevä yhteys kuolleisuuteen. Kuoleman todennäköisyys näyttää laskevan, kun TAS-summan arvo on luvun 40 läheisyydessä ja kasvavan, kun summan arvo on luvun 50 läheisyydessä. Kuvan perusteella ei kuitenkaan voida tehdä päätelmiä aleksitymian vaikutuksesta kuolleisuuteen, koska aleksityymisyys ilmenee vasta summan ollessa suurempi kuin 60.

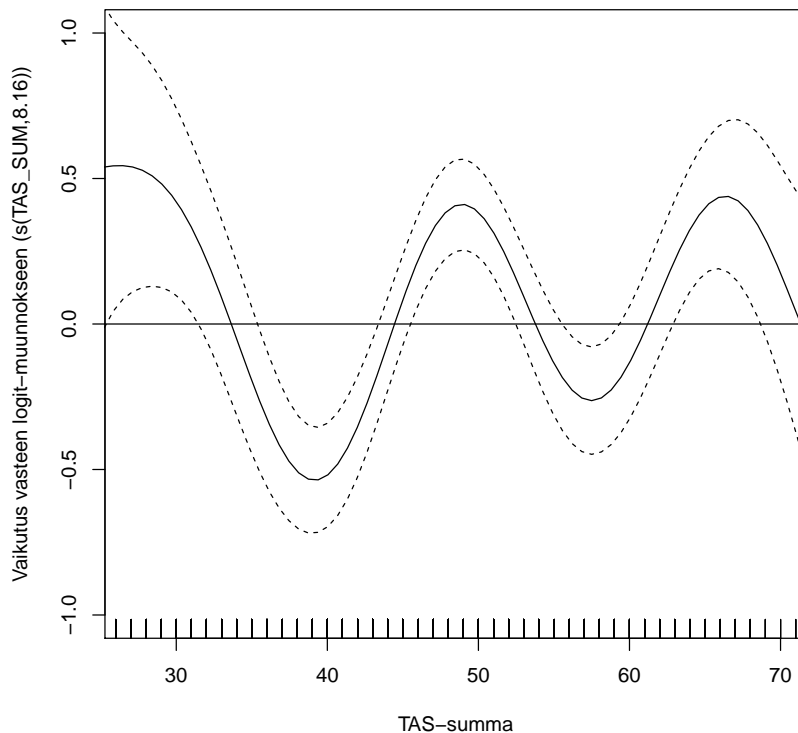


Kuva 4.1. TAS-summan yhteys yleiseen kuolleisuuteen.

Jatketaan tarkastelua rajaamalla kuolemat vain niihin, joiden syyksi on määritelty sydän- ja verisuonitaudit. Malli on kaavan (4.1) mukainen, koska tasoitettaviin termeihin ei tarvinnut tehdä muutoksia. Devianssista saadaan nyt selitettyä 33,1 % eli hieman enemmän kuin yleisen kuolleisuuden mallissa. Lineaarisesti mallinnettavista selittäjistä ainoastaan depression ($p=0,14$) ja alkoholin kulutus ($p=0,29$) eivät ole tilastollisesti merkitseviä (liite B, taulukko 2). Tasoittavien funktioiden avulla mallinnettavista selittäjistä merkitseviä ovat kaikki muut paitsi koulutus ($p=0,17$), jonka efektiiviset vapausasteet olisivat 3,06. TAS-summalla on 8,16 efektiivistä vapausastetta, iällä 8,30 ja painoindeksillä 7,44.

Kuvasta 4.2 nähdään, miten TAS-summa vaikuttaa sydän- ja verisuonitautikuolleisuuteen. Käyrä aaltoilee samaan tapaan kuin edellisessäkin kuvassa, mutta sen huiput ovat nyt likimain TAS-summan arvojen 26, 50 ja 66 kohdilla. Käyrän luottamusväli ulottuu kokonaan nollan alapuolelle, kun TAS-summa

on väleillä (36–43) tai (55–60), ja vastaavasti nollan yläpuolella, kun summa on väleillä (25–31), (46–52) tai (63–68). Näillä alueilla yhteys kuolleisuuteen on merkitsevä. Viimeinen merkitsevä alue on aleksityymisyyden rajan yläpuolella (>60), joten aleksityymialla näyttää olevan positiivinen yhteys sydän- ja verisuonitautikuolleisuuteen. Merkitsevä alue on kuitenkin hyvin pieni, eikä kata kaikkia aleksityymiä. Tuloksiin voi vaikuttaa kuolleiden pieni määrä. Seuranta-aikaa pidentämällä olisi mahdollista selvittää, voimistuuko vai häviääkö nyt havaittu yhteys suuremmalla kuolleiden määrällä.

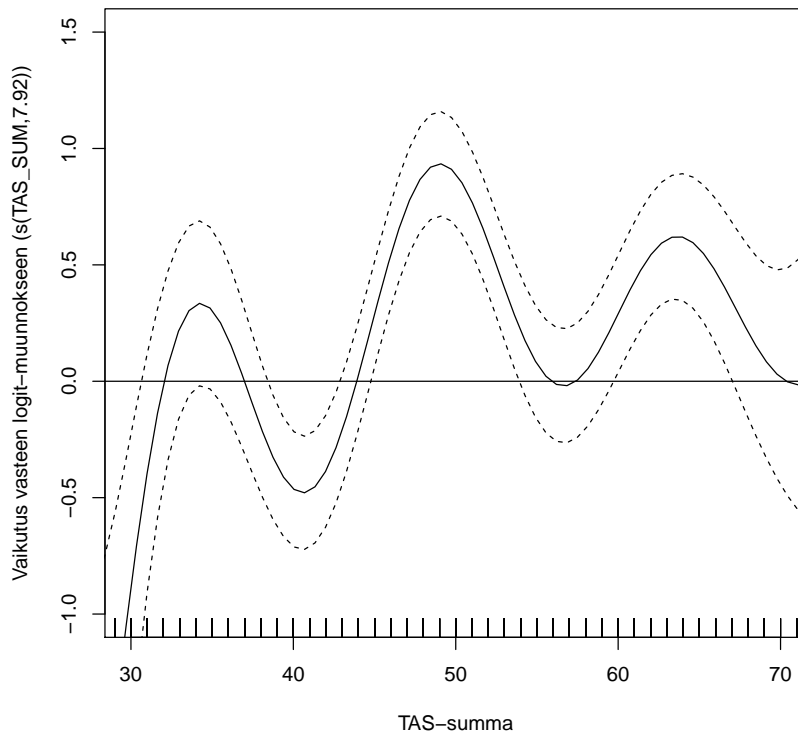


Kuva 4.2. TAS-summan yhteys sydän- ja verisuonitautikuolleisuuteen.

Seuraavaksi tarkastellaan vielä tarkemmin rajattua kuolleiden joukkoa, kun selitettäväksi muuttujaksi valitaan kuolleisuus verenpaine- tai iskeemisiin sydänsairauksiin. Alkuperäiseen malliin ei tehty muutoksia. Tulokset on esitetty liitteen B taulukossa 3. Tämän mallin avulla saadaan selitettyä 32,2 % devianssista. Kaikki selittävät muuttujat ovat alkoholin kulutusta ($p=0,29$) lukuun ottamatta tilastollisesti merkitseviä. TAS-summalla on 7,92 efektiivistä vapausastetta, iällä 8,71, koulutuksella 3,99 ja painoindeksillä 8,98.

Mallin antama tulos TAS-summan ja verenpaine- ja sydänsairauskuolleisuuden väliselle yhteydelle on kuvan 4.3 mukainen. Kuten aiemmissa malleissa, tässäkin käyrä aaltoilee nollan ympärillä. Käyrä on edelleen kolmihiippuinen siten, että huiput ovat TAS-summan arvojen 34, 49 ja 64 kohdilla. Käyrän luot-

tamuväli ulottuu kokonaan nollan alapuolelle, kun TAS-summa on alle 31 tai välillä (39–42). Näillä summan arvoilla kuolleisuuden todennäköisyys pienenee. Kun summa on väleillä (45–55) tai (60–68), luottamusväli on nollan yläpuolella. Silloin kuolleisuuden todennäköisyys kasvaa. Kuten edellisessä mallissa, tässäkin viimeinen merkitsevä alue ulottuu aleksityymisyyden rajan yläpuolelle (>60), joten aleksityymialla näyttää olevan positiivinen yhteys verenpaine- ja sydänsairauskuolleisuuteen. Merkitsevä alue ei tässäkään mallissa kuitenkaan kata kaikkia aleksityymiä.

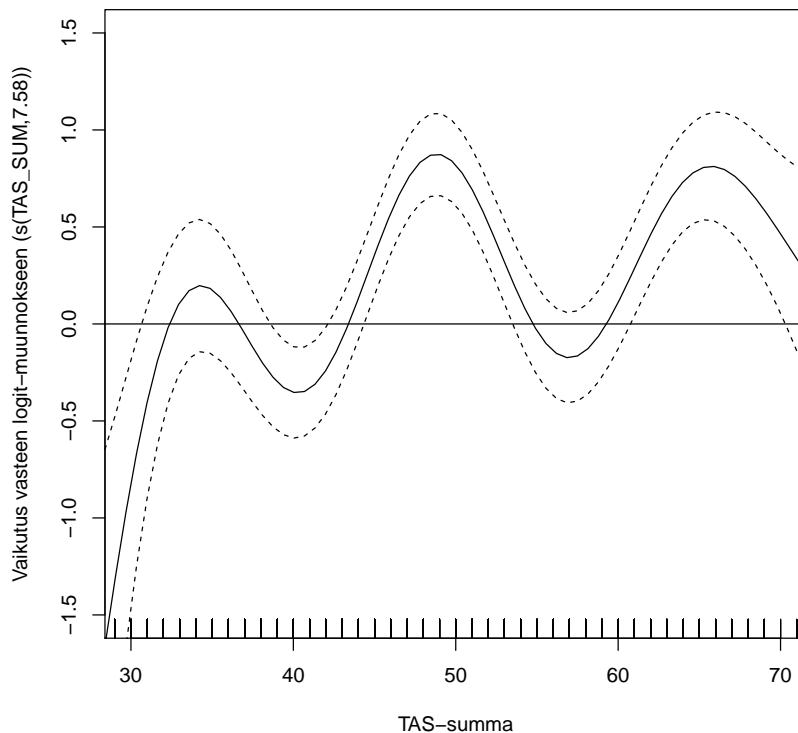


Kuva 4.3. TAS-summan yhteys verenpaine- tai iskeemiseen sydänsairauskuolleisuuteen.

Lopuksi tarkastellaan mallia, jossa edellisen mallin kuolleiden joukkoon otetaan lisäksi aivoinfarktiin kuolleet. Alkuperäiseen malliin ei tarvinnut tehdä muutoksia, joten kaikki neljä kvantitatiivista selittäjää lisätään malliin tasoittavien funktioiden avulla. Devianssista saadaan nyt selitettyä 33,6 %. Selittävistä muuttujista vain depressio ($p=0,11$) ja alkoholin kulutus ($p=0,17$) eivät ole tilastollisesti merkitseviä (ks. liite B, taulukko 4). Korkeimmat efektiiviset vapausasteet on iällä (8,64) ja toiseksi korkeimmat painoindeksillä (8,30). TAS-summalla vapausasteita on 7,58 ja koulutuksella 3,52.

Kuvassa 4.4 on mallin antama estimaatti TAS-summan tasoittavalle funktiolle. Käyrä on samaan tapaan aaltoileva ja kolmihuippuinen kuin edellisten-

kin mallien käyrät. Huiput saavutetaan TAS-summan arvoilla 34, 49 ja 66. Käyrä on summan pienimmillä arvoilla nollan alapuolella, kuten edellisessäkin mallissa. TAS-summalla on merkitsevä yhteys aivoinfarkti-, verenpaine- ja sydänsairauskuolleisuuteen, kun summa on pienempi kuin 31 tai väleillä (39–42), (44–54) ja (61–70). Ensimmäisillä kahdella summan arvoalueella vaikutus on kuolleisuutta vähentävä ja kahdella jälkimmäisellä kuolleisuutta lisäävä. Viimeinen merkitsevä alue on jälleen aleksityymisyyden rajan yläpuolella (>60), joten aleksitymiällä näyttää olevan positiivinen yhteys kuolleisuuteen. Kuitenkin on huomattava, ettei merkitsevä alue kata kaikkia aleksityymiä tässäkin mallissa.



Kuva 4.4. TAS-summan yhteys kuolleisuuteen, kun huomioidut kuolinsyyt ovat verenpaine- tai iskeemiset sydänsairaudet tai aivoinfarkti.

Seuraavaksi tarkastellaan kaavan (4.2) mukaisia interaktiomalleja. Interaktioiden tasoitteiden funktioiden estimoinnissa käytetään tensoritulotasoitteita thin plate -regressiosplinkannoilla. Yksittäisten selittäjien tasoitteet estimoidaan näissäkin malleissa thin plate -regressiospliniin avulla.

Jokainen malli on muodostettu siten, että edellä esitettyihin malleihin on lisätty yksi interaktiotermin kerrallaan. Näitä yhden interaktiotermin malleja verrattiin alkuperäiseen malliin, ja jos uusi malli oli alkuperäistä merkittävästi parempi, interaktiotermin pidettiin mallissa. Mallien vertailussa käytet-

tiin χ^2 -testiä. Yksittäisten interaktiitermien merkitsevyyksien testauksen jälkeen kaikki merkitsevät interaktiotermit yhdistettiin samaan malliin. Lopullisen mallin merkitsevyyttä tutkittiin samaan tapaan kuin yhden interaktiotermin malleja. Mallien testauksen tulokset on esitetty liitteessä C. Malleja paranneltiin alkuperäisten tapaan poistamalla tarvittaessa termien tasoituksia. Interaktiotermejä lisättiin vain TAS-summan ja muiden tasoitettavien selittäjien välille.

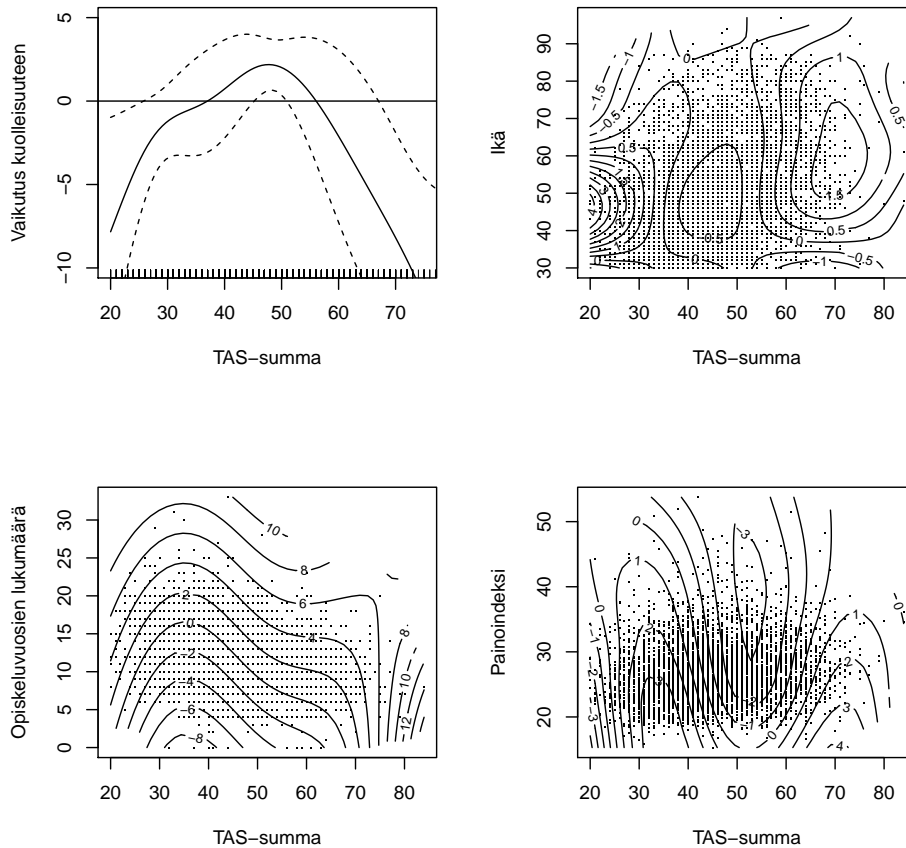
Tarkastellaan aluksi mallia, jossa selitettävänä muuttujana on yleinen kuolleisuus. Yksittäisten interaktioiden lisääminen paransi mallia tilastollisesti merkitsevästi ($p < 0,001$). Kun malliin lisättiin kaikki interaktiot, huomattiin, että koulutuksen mallintaminen tasoitettavien funktioiden avulla ei ole tarpeen ($edf=1.00$).

Lopullisessa mallissa tasoitettuja termejä ovat siis TAS-summa, ikä, painoindeksi ja kaikki kolme interaktiota. Interaktiomalli on alkuperäistä tilastollisesti merkitsevästi parempi ($p < 0,001$). Devianssista saadaan selitettyä 30,9 %. Kaikki muut selittävät muuttujat ovat tilastollisesti merkitseviä, paitsi koulutus ($p=0,053$), jonka p-arvo on kuitenkin hyvin lähellä viiden prosentin riskitasoa (ks. liite B, taulukko 5). TAS-summalla on 6,84 efektiivistä vapausastetta, iällä 2,41 ja painoindeksillä 7,23. TAS-summan ja iän välisen interaktion efektiiviset vapausasteet ovat korkeimmat ($edf=11,995$). TAS-summan interaktiolla koulutuksen kanssa on vapausasteita on 5,96 ja painoindeksin kanssa 7,42.

Kuvasta 4.5 nähdään, miten TAS-summa ja sen interaktiot iän, koulutuksen ja painoindeksin kanssa vaikuttavat yleisen kuolleisuuden logit-muunnokseen. Ylärivillä ensimmäisessä kuvassa on esitetty pelkän TAS-summan, ja toisessa kuvassa TAS-summan ja iän välisen interaktion vaikutus kuolleisuuteen. Alarivillä on TAS-summan interaktio ensimmäisessä kuvassa koulutuksen ja toisessa painoindeksin suhteen.

Kahden muuttujan välisen interaktion vaikutusta kuolleisuuteen tarkastellaan ”contour”-kuvaajien (*contour plots*) avulla. Jos muuttujien välillä on interaktiota, kuvaajan viivat eivät ole suoria. Silloin toinen muuttuja vaikuttaa siihen, minkälainen vaikutus toisella muuttujalla on vasteeseen. Kuvaajissa näkyvien viivojen kohdalla olevat numerot kuvaavat yhteyden suuntaa ja voimakkuutta. Negatiiviset arvot tarkoittavat, että interaktiolla on kuolleisuutta vähentävä vaikutus ja positiiviset päin vastoin, että interaktiolla on kuolleisuutta lisäävä vaikutus. Mitä suurempi luku on, sitä voimakkaampi on interaktion vaikutus on vasteeseen.

Pelkän TAS-summan vaikutusta esittävässä kuvassa käyrä ei aaltoile alkuperäisten mallien kuvaajien tapaan. Käyrä on yksihiippuinen ja sen hännät ovat nollan alapuolella. Käyrän luottamusväli ulottuu kokonaan nollan alapuolelle, kun TAS-summa on pienempi tai yhtä suuri kuin 25 ja suurempi tai yhtä suuri kuin 67. Luottamusväli ylittää nollan yläpuolelle summan ollessa välillä (45–51). Voimakkaan aleksityymisyyden raja (>60) ylittyy toisella nollan alapuolelle jäävällä merkitsevällä alueella, jolla summa on suurempi tai yhtä suuri kuin 67. Aleksitymiällä näyttäisi siis olevan kuolleisuutta alentava omavaikutus, kun iän, koulutuksen ja painoindeksin yhdysvaikutukset on huomioitu mallis-



Kuva 4.5. TAS-summan ja interaktioiden yhteys yleiseen kuolleisuuteen.

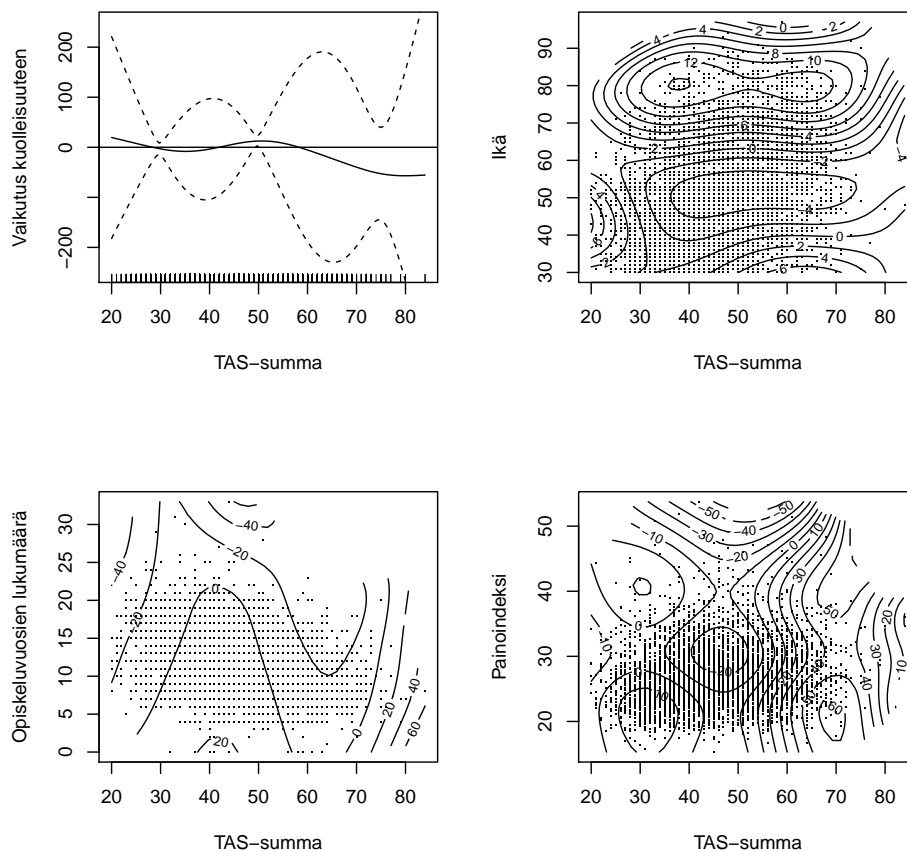
sa. Aleksitymia vaikuttaa kuolleisuuteen kuitenkin myös interaktiotermeissä, joten seuraavaksi tarkastellaan sen vaikutusta interaktiokuvaajien perusteella.

Iän ja TAS-summan välisen interaktion kuvaajasta nähdään, että voimakkaasti aleksityymisten (TAS-summa > 60) joukossa vaikutus kuolleisuuteen on positiivinen, kun ikää on yli 40 vuotta. Opiskeluvuosien määrän ja TAS-summan interaktiolla näyttää olevan kuolleisuutta lisäävä vaikutus, kun tarkastellaan voimakkaasti aleksityymisten joukkoa. Myös painoindeksin ja TAS-summan välisen interaktion kuvaaja viittaa kuolleisuuden todennäköisyyden kasvuun voimakkaasti aleksityymisten joukossa. Pienimmillä painoindeksin arvoilla ja suurilla TAS-summan arvoilla positiivinen yhteys on voimakkaimmillaan. Kuvaajien perusteella aleksitymian vaikutus kuolleisuuteen on positiivinen kaikkien interaktioiden osalta, mutta negatiivinen pelkän TAS-summan osalta. Kokonaisuudessaan aleksitymiällä näyttäisi siis olevan yleistä kuolleisuutta lisäävä vaikutus.

Jatketaan tarkastelua rajaamalla kuolemat vain niihin, joiden syyksi on määritetty sydän- ja verisuonitaudit. Jokainen yksittäinen interaktiotermi pa-

ransi mallia merkitsevästi ($p < 0,001$), joten ne kaikki lisättiin malliin. Koulutus on siirretty lopullisessa mallissa lineaarisesti selittäjäksi, koska sen mallintaminen tasoittavien funktioiden avulla ei ollut järkevää ($edf = 1,00$). Interaktioiden lisääminen paransi alkuperäistä mallia merkitsevästi ($p < 0,001$). Mallin tulokset on esitetty liitteen B taulukossa 6 ja mallien vertailun tulokset liitteessä C. Devianssista saadaan nyt selitettyä 37,7 %. Mallin parametrit ovat koulutusta ($p = 0,897$) ja alkoholin kulutusta ($p = 0,23$) lukuun ottamatta merkitseviä. TAS-summalla on 7,97 efektiivistä vapausastetta, iällä 8,22 ja painoindeksillä 8,998. TAS-summan interaktiolla iän kanssa on 13,86 efektiivistä vapausastetta, koulutuksen kanssa 17,64 ja painoindeksin kanssa 15,39.

Kuvasta 4.6 nähdään, miten TAS-summa vaikuttaa sydän- ja verisuonitautikuolleisuuteen. Käyrällä on loivat kaarteet likimain TAS-summan arvojen 34, 52 ja 76 kohdilla. Käyrän luottamusväli sisältää nollan kaikilla TAS-summan arvoilla, joten aleksitymiällä ei ole merkitsevää omavaikutusta sydän- ja verisuonitautikuolleisuuteen, kun iän, koulutuksen ja painoindeksin interaktiot TAS-summan kanssa on huomioitu.



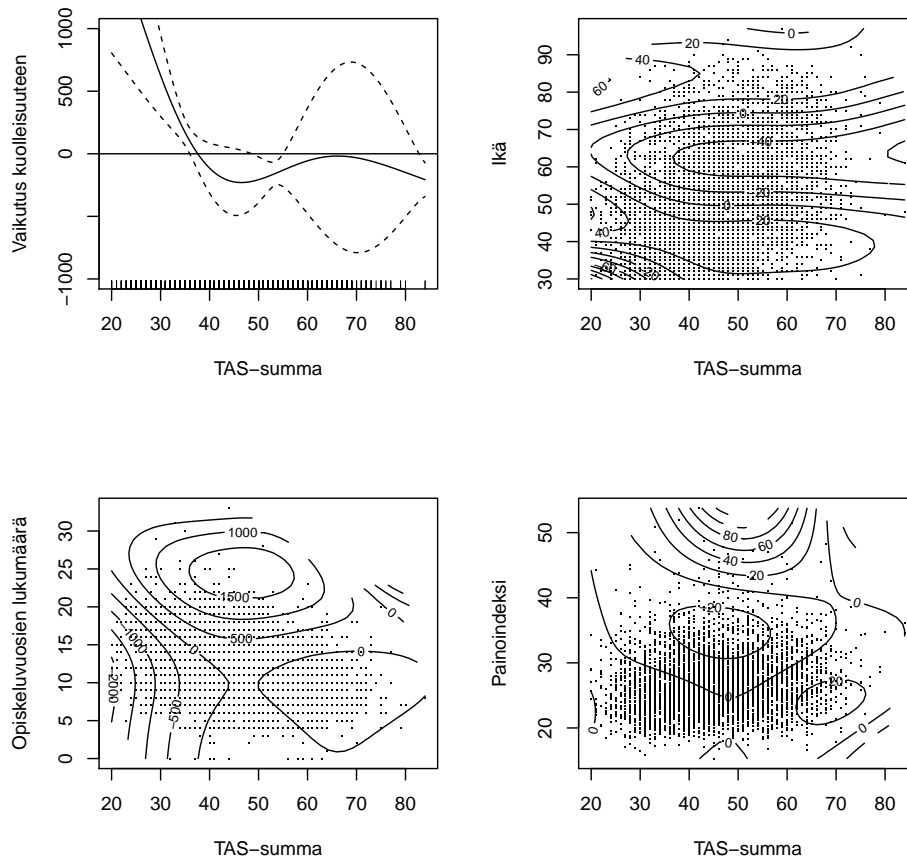
Kuva 4.6. TAS-summan ja interaktioiden yhteys sydän- ja verisuonitautikuolleisuuteen.

Jatketaan tarkastelua interaktiotermin kuvaajien avulla. Iän ja TAS-summan välisen interaktion kuvaajasta (toinen ylärivin kuvaaja) nähdään, että voimakkaasti aleksityymisten (TAS-summa > 60) joukossa interaktion vaikutus kuolleisuuteen on positiivinen, kun ikää on yli 60 tai alle 40 vuotta. Voimakas aleksityymisyys näyttäisi lisäävän kuolleisuuden todennäköisyyttä yhtä paljon 30-vuotiailla kuin lähes 70-vuotiailla. Opiskeluvuosien määrän ja TAS-summan interaktiolla näyttää olevan kuolleisuutta lisäävä vaikutus, kun tarkastellaan voimakkaasti aleksityymisten joukossa kaikkein suurimpia pistemääriä (TAS-summa > 70). Myös painoindeksin ja TAS-summan välisen interaktion kuvaaja viittaa kuolleisuuden todennäköisyyden kasvuun aleksityymisten joukossa. Positiivinen yhteys on voimakkaimmillaan, kun TAS-summa on 70 pisteen ja painoindeksi luvun 20 kohdalla. Aleksityymian vaikutus kuolleisuuteen näyttää olevan positiivinen kaikkien interaktiotermin osalta, mutta pelkän TAS-summan osalta yhteys ei ole merkitsevä. Kokonaisuudessaan aleksityymialla näyttäisi siis olevan sydän- ja verisuonitautikuolleisuutta lisäävä vaikutus.

Rajataan seuraavassa mallissa kuolemat koskemaan vain kuolleisuutta verenpaine- tai iskeemisiin sydänsairauksiin. Yksittäiset interaktiotermit olivat tilastollisesti merkitseviä ($p < 0,001$), joten lopullinen malli on kaavan (4.2) mukainen ja sen tulokset on esitetty liitteen B taulukossa 7. Tasoitettaviin termeihin ei tarvinnut tehdä muutoksia. Interaktiomallin devianssista saadaan selitettyä 38,8 % eli 6,6 %-yksikköä enemmän kuin tavallisessa mallissa. Ero on tilastollisesti merkitsevä ($p < 0,001$). Selittävästä muuttujista ainoastaan alkoholin kulutus ($p = 0,28$) ei ole tilastollisesti merkitsevä. TAS-summalla on 8,29 efektiivistä vapausastetta, iällä 9,00, koulutuksella 8,88 ja painoindeksillä 8,98. TAS-summan interaktiolla iän kanssa on 15,40 efektiivistä vapausastetta, koulutuksen kanssa 15,89 ja painoindeksin kanssa 13,01.

Kuvassa 4.7 on esitetty kuvaajat TAS-summan ja interaktiotermin vaikutuksesta verenpaine- tai sydänsairauskuolleisuuteen. TAS-summan omavaikutus kuolleisuuteen on esitetty ensimmäisessä yläkuvassa. Käyrällä on loivat kaarteet TAS-summan arvojen 45 ja 65 kohdilla. Käyrän luottamusväli ulottuu kokonaan nollan yläpuolelle, kun TAS-summa on korkeintaan 36. Summan ollessa 50 pisteen kohdilla, luottamusväli on pienellä alueella nollan alapuolella. Muilla summan arvoilla nolla kuuluu luottamusvälille, eikä yhteys ole tilastollisesti merkitsevä. TAS-summan pienimmillä arvoilla näyttäisi siis olevan positiivinen omavaikutus kuolleisuuden logit-muunnokseen, kun mallissa on huomioitu iän, koulutuksen ja painoindeksin interaktiot TAS-summan kanssa. Koska merkitsevä alue ei ole aleksityymisyyden rajan yläpuolella, aleksityymialla ei näyttäisi olevan yhteyttä kuolleisuuteen.

Iän ja TAS-summan välisen interaktion kuvaajasta nähdään, että alle 50-vuotiaiden ja yli 75-vuotiaiden osalta aleksityymialla on positiivinen yhteys kuolleisuuteen. Alle 40 pisteen TAS-summalla näyttäisi olevan kuolleisuutta lisäävä vaikutus yli 40-vuotiaiden joukossa. Kun tarkastellaan voimakkaasti aleksityymisten joukkoa (TAS-summa > 60), opiskeluvuosien määrän ja TAS-summan interaktiolla ei näytä olevan suurta vaikutusta kuolleisuuteen. Pienillä TAS-summan arvoilla näyttäisi olevan kuolleisuutta vähentävä vaikutus lähes kaikil-



Kuva 4.7. TAS-summan ja interaktioiden yhteys verenpaine- tai iskeemiseen sydänsairauskuolleisuuteen.

la opiskeluvuosien määrillä. Painoindeksin ja TAS-summan välisen interaktion kuvaaja viittaa kuolleisuuden todennäköisyyden kasvuun voimakkaasti aleksityymisten joukossa vain painoindeksin ollessa arvojen 20 ja 30 välillä. Pienillä TAS-summan arvoilla ei näytä olevan yhteyttä kuolleisuuteen millään painoindeksin arvoilla.

Kuvaajien perusteella aleksitymian vaikutus verenpaine- tai sydänsairauskuolleisuuteen ei näyttäisi olevan kovin voimakas, mutta iän ja TAS-summan välisellä interaktiolla on positiivinen yhteys kuolleisuuteen alle 50- ja yli 75-vuotiaiden osalta. Koulutuksen ja painoindeksin väliset interaktiot TAS-summan kanssa paljastivat pienimmillä summan arvoilla negatiivisen yhteyden kuolleisuuteen. Tämä siis kumoaa osaltaan TAS-summan omavaikutuksella havaittua positiivista yhteyttä kuolleisuuteen. Näyttäisi siis siltä, ettei TAS-summalla eikä aleksitymiällä ole kuvaajien perusteella kovin voimakasta yhteyttä verenpaine- tai iskeemiseen sydänsairauskuolleisuuteen.

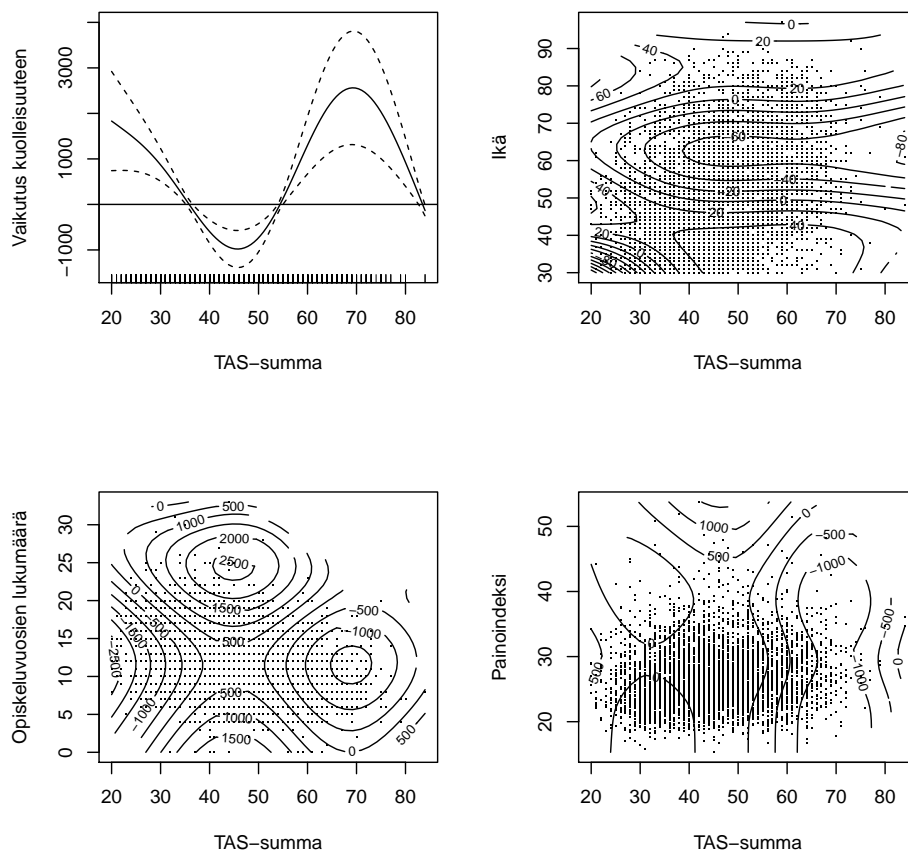
Lopuksi tarkastellaan mallia, jossa edellisessä mallissa tarkasteltuihin kuolemiin on lisätty aivoinfarktiin kuolleet. Kaikki yksittäiset interaktiotermit pa-

ransivat mallia tilastollisesti merkitsevästi (iän ja koulutuksen $p < 0,001$ ja painoindeksin $p = 0,001$), joten ne lisättiin malliin. Lopullinen malli on kaavan (4.2) mukainen, koska tasoitettaviin termeihin ei tarvinnut tehdä muutoksia. Mallin tulokset on esitetty liitteen B taulukossa 8 ja mallien vertailun tulokset liitteessä C. Devianssista saadaan nyt selitettyä 39,4 % ja ero tavalliseen malliin on tilastollisesti merkitsevä ($p < 0,001$). Alkoholin kulutusta ($p = 0,07$) lukuun ottamatta termit ovat tilastollisesti merkitseviä. TAS-summalla on 9,00 efektiivistä vapausastetta, iällä 8,90, koulutuksella 8,92 ja painoindeksillä 8,94. TAS-summan interaktiolla iän kanssa on 15,88 efektiivistä vapausastetta, koulutuksen kanssa 15,56 ja painoindeksin kanssa 15,98.

TAS-summan ja interaktiotermien yhteyttä aivoinfarkti-, verenpaine- tai sydänsairauskuolleisuuteen tarkastellaan kuvassa 4.8. TAS-summan kuvaajassa käyrä aaltoilee nollan molemmiin puolin, ja sen huiput ovat TAS-summan arvojen 20 ja 69 kohdilla. Käyrän luottamusväli on nollan alapuolella, kun TAS-summa on välillä (37–54) ja vastaavasti nollan yläpuolella, kun summa on alle 35 tai välillä (55–83). Näillä alueilla yhteys kuolleisuuteen on merkitsevä. Viimeinen merkitsevä alue ulottuu aleksityymisyyden rajan yläpuolelle (> 61). Aleksitymiällä näyttää siis olevan positiivinen yhteys aivoinfarkti-, verenpaine tai sydänsairauksien aiheuttamaan kuolleisuuteen, kun iän, koulutuksen ja painoindeksin interaktiot TAS-summan kanssa on huomioitu.

Iän ja TAS-summan välisen interaktion kuvaajasta nähdään, että aleksityymisten joukossa vaikutus kuolleisuuteen on positiivinen, kun ikää on alle 50 tai yli 75 vuotta. Opiskeluvuosien määrän ja TAS-summan interaktiolla näyttää olevan kuolleisuutta alentava vaikutus, kun tarkastellaan voimakkaasti aleksityymisten joukkoa (TAS-summa > 60). Ainoastaan suurimmilla TAS-summan arvoilla (> 80) interaktion vaikutus on positiivinen. Myös painoindeksin ja TAS-summan välisen interaktion kuvaaja viittaa kuolleisuuden todennäköisyyden laskemiseen voimakkaasti aleksityymisten joukossa.

Kuvaajien perusteella aleksityymien vaikutus kuolleisuuteen on positiivinen TAS-summan omavaikutuksen ja iän interaktion osalta, mutta negatiivinen muiden interaktioiden osalta. Kokonaisuudessaan aleksitymiällä näyttäisi olevan yhteys aivoinfarkti-, verenpaine- tai sydänsairauskuolleisuuteen, mutta sen voimakkuutta heikentää se, että osa termeistä vaikuttaa kuolleisuuteen positiivisesti ja osa negatiivisesti.



Kuva 4.8. TAS-summan yhteys aivoinfarkti-, verenpaine- tai sydänsairauskuolleisuuteen.

5 Yhteenveto

Tutkielman tavoitteena oli selvittää, onko aleksitymialla yhteyttä kuolleisuuteen. Tutkimusaineistona käytettiin Terveys 2000 -tutkimuksessa kerättyä aineistoa. Kuolleisuuden mallintamiseen käytettiin yleistettyjä additiivisia malleja siten, että tasoittavien funktioiden estimoinnissa kantana käytettiin thin plate -splinejä, ja interaktioiden tapauksessa lisäksi tensoritulotasoitetta. Yleistettyjen additiivisten mallien etuna yleistettyihin lineaarisiin malleihin nähden on monimutkaisempien mallien tehokkaampi muodostaminen, mutta kääntöpuolena monimutkaisempien mallien tulkinta voi olla hankalampaa.

Tutkielmassa tarkasteltiin aleksitymian yhteyttä kuolleisuuteen neljällä eri kuolleiden ryhmällä. Ensimmäisessä mallissa tarkasteltiin kaikkia kuolleita. Toisessa mallissa vasteena oli kuolleisuus sydän- ja verisuonitauteihin, kolmannessa verenpaine- tai sydänsairauksiin ja neljännessä aivoinfarkti-, verenpaine- tai sydänsairauksiin. Näiden varsinaisten mallien lisäksi tutkittiin interaktiomallien avulla aleksitymian mahdollista yhdysvaikutusta iän, koulutuksen ja painoindeksin suhteen. Interaktiomalleissa tarkasteltiin samoja neljää eri kuolleisuuden ryhmää kuin tavallisissakin malleissa.

Kun selitettävänä muuttujana oli kuolleisuus ilman kuolinsyyrajauksia, aleksitymialla ei havaittu yhteyttä kuolleisuuteen. Pieniä merkitseviä alueita havaittiin TAS-summan arvojen 40 ja 50 läheisyydessä. Interaktioiden lisäämisen jälkeen yhteys kuolleisuuteen voimistui. TAS-summan oma vaikutuksen tapauksessa merkitseviä yhteyksiä oli nyt havaittavissa, kun TAS-summa oli korkeintaan 25, vähintään 67 tai välillä (45–51). Aleksityymisyyden alueella summan vaikutus kuolleisuuteen oli yllättäen negatiivinen. Kuitenkin TAS-summan ja iän, koulutuksen sekä painoindeksin välisten interaktioiden kohdalla havaittiin, että aleksitymialla on positiivinen yhteys kuolleisuuteen. Interaktiomallin perusteella aleksitymialla näyttäisi siis olevan yhteys kuolleisuuteen siten, että kuolleisuuden todennäköisyys kasvaa TAS-summan ja iän, koulutuksen sekä painoindeksin yhdysvaikutusten osalta ja laskee TAS-summan oma vaikutuksen osalta.

Sydän- ja verisuonitautikuolleisuutta tarkasteltaessa havaittiin, että aleksitymialla saattaa olla yhteys kuolleisuuteen. Yhteys oli positiivinen, kun TAS-summan arvot olivat väleillä (25–31), (46–52) tai (63–68). Viimeisellä välillä summa ylittää aleksityymisyyden rajan, joten aleksitymialla näyttäisi olevan lievä kuolleisuutta lisäävä vaikutus. Interaktiomallissa aleksityymisyyden oma vaikutus kuolleisuuteen ei ollut tilastollisesti merkitsevä, mutta interaktioiden iän, koulutuksen ja painoindeksin kanssa paljastivat positiivisen yhteyden kuolleisuuteen.

suuteen.

Kun vasteeksi valittiin verenpaine- ja iskeemiset sydänsairauskuolemat, aleksitymiällä näytti olevan yhteys kuolleisuuteen. Yhteys kuolleisuuteen oli merkitsevä, kun TAS-summa oli pienempi kuin 31 tai väleillä (39–42), (45–55) tai (60–68). Koska viimeisellä merkitsevällä välillä ylitetään aleksityymisyyden katkaisupiste (>60), näyttäisi aleksitymiällä olevan tässäkin tapauksessa lievä kuolleisuutta lisäävä vaikutus. Interaktiomallissa aleksitymiällä ei ollut havaittavissa kovin voimakasta yhteyttä kuolleisuuteen. Ainoastaan iän ja aleksityymisyyden välisellä interaktiolla oli positiivinen vaikutus kuolleisuuteen alle 50- ja yli 75-vuotiaiden keskuudessa.

Viimeiseksi tutkittiin mallia, jossa verenpaine- ja iskeemisten sydänsairauksien lisäksi kuolinsyistä huomioitiin aivoinfarktit. Nyt yhteys oli merkitsevä TAS-summan ollessa alle 31 tai arvoalueilla (39–42), (44–54) ja (61–70). Aleksityymisyyden rajan ylittävällä merkitsevällä alueella yhteys kuolleisuuteen on positiivinen, joten aleksitymiällä näyttäisi olevan tässäkin mallissa kuolleisuutta lisäävä vaikutus. Interaktioiden lisääminen malliin voimisti tätä yhteyttä omavaikutuksen osalta. Interaktiomallin merkitsevät alueet havaittiin TAS-summan väleillä (37–54) tai (55–83). Viimeinen merkitsevä alue kattaa jo lähes kaikki aineiston havaitut aleksityymiset, joten aleksitymiällä näyttäisi olevan positiivinen omavaikutus aivoinfarkti-, verenpaine- ja sydänsairauskuolleisuuteen, kun iän, koulutuksen ja painoindeksin yhdysvaikutustermit TAS-summan kanssa oli huomioitu mallissa. Aleksityymisyyden ja iän välisellä interaktiolla oli myös havaittavissa kuolleisuutta lisäävä vaikutus, mutta muiden interaktioiden osalta yhteys kuolleisuuteen oli negatiivinen. Kokonaisuudessaan aleksityymisyydellä näyttäisi olevan yhteys kuolleisuuteen, mutta osalla termeistä yhteys oli positiivinen ja osalla negatiivinen.

Kuolleisuustietoja oli tutkimushetkellä saatavilla vain vuoden 2008 loppuun, joten lyhyt seuranta-aika voi vaikuttaa tuloksiin. Analyysissä ei huomioitu otantamenetelmän tai väestöpainojen vaikutusta. Vaikka otantamenetelmän vaikutus ei aineiston esittelyosan tunnuslukuvertailujen perusteella näyttänyt olevan kovinkaan suuri, sillä ja painoilla voi kuitenkin olla jonkin verran vaikutusta tuloksiin.

Aleksitymian ja taustamuuttujien väliset syy-seurausyhteydet vaativat vielä suurelta osin lisätutkimusta. On kuitenkin hyvä pitää mielessä, että aleksitymia on voimakkaasti yhteydessä taustamuuttujiin. Vaikka aleksitymian ja kuolleisuuden välillä ei löytyisi suoraa merkitsevää yhteyttä, aleksityymisyys voi silti vaikuttaa kuolleisuuteen taustamuuttujien kautta.

Lähteet

- Agresti, A. (2007), "An introduction to Categorical Data Analysis", New Jersey: John Wiley & Sons.
- Chatzi, L., Bitsios, P., Solidaki, E., Christou, I., Kyrlaki, E., Sfakianaki, M., Kogevinas, M., Kefalogiannis, N. & Pappas, A. (2009), "Type 1 diabetes is associated with alexithymia in nondepressed, non-mentally ill diabetic patients: A case-control study", *Journal of Psychosomatic Research*, 67, 307–313.
- Eubank, R. L. (1999), "Nonparametric regression and spline smoothing", *STATISTICS: textbooks and monographs*, New York: Marcel Dekker.
- Guisan, A., Edwards, T. C. & Hastie, T. J. (2002), "Generalized linear and generalized additive models in studies of species distributions: setting the scene", *Ecological Modelling*, 157, 89–100.
- Hastie, T. J. & Tibshirani, R. J. (1990), "Generalized Additive Models", *Monographs on Statistics and Applied Probability*, London: Chapman & Hall.
- Heistaro S., toim. (2005), "Menetelmäraportti. Terveys 2000 -tutkimuksen toteutus, aineisto ja menetelmät", *Kansanterveyslaitoksen julkaisuja B6/2005*, Helsinki. Saatavilla Internetistä: <http://www.terveys2000.fi/julkaisut/2005b6.pdf>.
- Helmers, K. F. & Mente, A. (1999), "Alexithymia And Health Behaviors In Healthy Male Volunteers", *Psychomatic Medicine*, 47, 635–645.
- Isotalo, J. (2009), "Yleistetyt lineaariset mallit I", *Luentomoniste*.
- Kauhanen, J., Kaplan, G. A., Cohen, R. D., Julkunen, J. & Salonen, J. T. (1996), "Alexithymia and risk of death in middle-aged men", *Journal of Psychosomatic Research*, 41, 541–549.
- Kojima, M., Hayano, J., Suzuki, S., Seno, H., Kasuga, H., Takahashi, H., Toriyama, T., Kawahara, H. & Furukawa, T. A. (2010), "Depression, Alexithymia and Long-Term Mortality in Chronic Hemodialysis Patients", *Psychotherapy and Psychosomatics*, 79, 303–311.
- Laiho J., Nieminen T., toim. (2004), "Terveys 2000 -tutkimus. Aikuisväestön haastatteluaikaväestön tilastollinen laatu. Otanta-asetelma, tiedonkeruu, vastauskato ja estimointi- ja analysointiasetelma", *Tilastokeskus, Tutkimuksia 239*, Helsinki.
- Lumley, M. A., Stettner, L. & Wehmer, F. (1996), "How Are Alexithymia And Physical Illness Linked? A Review And Critique Of Pathways", *Journal of Psychosomatic Research*, 41, 505–518.
- Lumley, M. A., Asselin, L. A. & Norman, S. (1997), "Alexithymia in Chronic Pain Patients", *Comprehensive Psychiatry*, 38, 160–165.
- Marchesi, C., Brusamonti, E. & Maggini, C. (2000), "Are alexithymia, depression,

- and anxiety distinct constructs in affective disorders?”, *Journal of Psychosomatic Research*, 49, 43–49.
- Mattila, A. (2009), ”Alexithymia in Finnish General Population”, *Väitöskirja*, Terveystieteen laitos, Tampere: Tampereen Yliopistopaino, Acta Universitatis Tamperensis 1377.
- Mustajoki, P. (2011), ”Painoindeksi (BMI)”, *Lääkärikirja Duodecim*. Saatavilla Internetistä: http://www.terveyskirjasto.fi/terveyskirjasto/tk.koti?p_artikkeli=dlk01001 (30.5.2012).
- Nummi, T. (2008), ”Ei-parametrinen regressio”, *Luentomoniste*. Saatavilla Internetistä: <http://www.sis.uta.fi/tilasto/eparam/Moniste/> (14.10.2012).
- Ruppert, D., Wand, M. P. & Carroll, R. J. (2003), ”Semiparametric Regression”, *Cambridge Series in Statistical and Probabilistic Mathematics*, New York: Cambridge University Press.
- Serrano, J., Plaza, V., Sureda, B., de Pablo, J., Picado, C., Bardag, S., Lamela, J. & Sanchis, J. (2006), ”Alexithymia: a relevant psychological variable in near-fatal asthma”, *European Respiratory Journal*, 28, 296–302.
- THL - Terveystieteiden ja hyvinvoinnin laitos (2009), ”Terveys 2000”, <http://www.terveys2000.fi/ohje.html> (11.5.2012).
- Tolmunen, T., Lehto, S. M., Heliste, M., Kurl, S. & Kauhanen, J. (2010), ”Alexithymia Is Associated With Increased Cardiovascular Mortality in Middle-Aged Finnish Men”, *Psychomatic Medicine*, 72, 187–191.
- Wood, S. N. (2006), ”Generalized Additive Models: An Introduction with R”, *Texts in Statistical Science*, Boca Raton: Chapman & Hall/CRC.
- (2012), ”GAMs with GCV/AIC/REML smoothness estimation and GAMMs by PQL”, *R Help*. Saatavilla Internetistä: cran.r-project.org/web/packages/mgcv/mgcv.pdf (14.10.2012).

Liite A: TAS-20-kyselylomake

TUNNE-ELÄMÄKYSELY

Toronto Alexithymia Scale (TAS-20)

Ympyröikää seuraavien toteamusten kohdalla jokin numeroista yhdestä viiteen sen mukaan, miten hyvin kukin toteamus sopii kuvaamaan Teitä.

	Ei lainkaan pidä paikkaansa	Ei juuri pidä paik- kaansa	En osaa sanoa	Pitää melko lailla paikkansa	Pitää täysin paikkansa
1. Olen usein epävarma siitä, mitä milloinkin tunnen.	1	2	3	4	5
2. Minun on vaikea löytää oikeita sanoja kuvaakseni tunteitani.	1	2	3	4	5
3. Minulla on fyysisiä tuntemuksia, joita lääkäritkään eivät ymmärrä.	1	2	3	4	5
4. Minun on helppo kuvailla tunteitani.	1	2	3	4	5
5. Mieluummin erittelen ja tutkin ongelmia kuin vain kuvailen niitä.	1	2	3	4	5
6. Kun olen poissa tolaltani, en tiedä olenko surullinen, peloissani vai vihainen.	1	2	3	4	5
7. Olen usein ymmälläni kehoni tuntemuksista.	1	2	3	4	5
8. Annan mieluummin asioiden mennä omalla painollaan kuin mietin mistä ne oikein johtuvat.	1	2	3	4	5
9. Minulla on tunteita, joita en täysin pysty tunnistamaan.	1	2	3	4	5
10. On erityisen tärkeää olla kosketuksissa tunteisiinsa.	1	2	3	4	5
11. Minun on vaikea kuvailla tunteita, joita toiset ihmiset minussa herättävät.	1	2	3	4	5
12. Ihmiset ovat kehottaneet minua kertomaan enemmän tunteistani.	1	2	3	4	5
13. En tiedä, mitä sisimmässäni oikein tapahtuu.	1	2	3	4	5
14. En aina tiedä, miksi olen vihainen.	1	2	3	4	5
15. Mieluummin puhun ihmisten kanssa heidän päivittäisistä puuhistaan kuin heidän tunteistaan.	1	2	3	4	5
16. Katselen mieluummin kevyttä viihdettä kuin psykologisia näytelmiä.	1	2	3	4	5
17. Minun on vaikea paljastaa sisimpiä tuntojani edes läheisille ystäville.	1	2	3	4	5
18. Voin tuntea läheisyyttä toiseen ihmiseen, vaikka oltaisiin hiljaa.	1	2	3	4	5
19. Olen huomannut, että omien tunteiden kuunteleminen ja pohtiminen auttaa henkilökohtaisten ongelmien ratkaisemisessa.	1	2	3	4	5
20. Elokvista tai näytelmistä häviää nautinto, jos niistä yrittää etsiä syvällisiä merkityksiä.	1	2	3	4	5

Liite B: Mallien tuloksia

Taulukko 1. Malli 1: Yleinen kuolleisuus ilman interaktioita. Devianssista selitetty 29,4 %.

Parametriset termit			
Muuttujan nimi	Estimaatti	P-arvo	
Vakio	-4,202559	< 2e-16	***
Koulutus	-0,003054	0,791414	
Sukupuoli	0,879620	< 2e-16	***
Siviilisääty	0,522320	5,23e-10	***
Depressio	0,382522	0,017734	*
Ahdistus	0,595689	0,001835	**
Alkoholi	0,314559	0,089889	.
Tupakka	0,884683	< 2e-16	***
Liikunta	-0,312776	0,000213	***
Metabolinen	0,345976	0,000147	***
Tasoitettut termit			
Muuttujan nimi	Edf	P-arvo	
TAS-summa	7,267	0,00223	**
Ikä	2,546	< 2e-16	***
Painoindeksi	7,254	1,29e-05	***

(***) 0,001 (***) 0,01 (*) 0,05 (.) 0,1.

Taulukko 2. Malli 2: Sydän- ja verisuonitautikuolleisuus ilman interak-
tioita. Devianssista selitetty 33,1 %.

Parametriset termit			
Muuttujan nimi	Estimaatti	P-arvo	
Vakio	-5,78041	< 2e-16	***
Sukupuoli	1,16967	< 2e-16	***
Siviilisääty	0,48933	4,11e-08	***
Depressio	-0,29537	0,135542	
Ahdistus	1,11253	5,63e-08	***
Alkoholi	-0,28561	0,292645	
Tupakka	0,73981	1,62e-11	***
Liikunta	-0,68550	2,64e-15	***
Metabolinen	0,35381	0,000125	***
Tasoitettut termit			
Muuttujan nimi	Edf	P-arvo	
TAS-summa	8,162	7,25e-10	***
Ikä	8,300	< 2e-16	***
Koulutus	3,060	0,165872	
Painoindeksi	7,437	0,000177	***

(***) 0,001 (**) 0,01 (*) 0,05 (.) 0,1.

Taulukko 3. Malli 3: Verenpaine- ja sydänsairauskuolleisuus ilman interaktioita. Devianssista selitetty 32,2 %.

Parametriset termit			
Muuttujan nimi	Estimaatti	P-arvo	
Vakio	-17,75969	< 2e-16	***
Sukupuoli	1,15246	< 2e-16	***
Siviilisääty	0,48524	9,89e-08	***
Depressio	-0,75471	0,00101	**
Ahdistus	1,31282	2,54e-09	***
Alkoholi	-0,29628	0,28606	
Tupakka	0,65686	3,59e-09	***
Liikunta	-0,73130	< 2e-16	***
Metabolinen	0,60947	9,99e-11	***
Tasoitettut termit			
Muuttujan nimi	Edf	P-arvo	
TAS-summa	7,921	< 2e-16	***
Ikä	8,714	< 2e-16	***
Koulutus	3,987	5,76e-06	***
Painoindeksi	8,984	< 2e-16	***

(***) 0,001 (**) 0,01 (*) 0,05 (.) 0,1.

Taulukko 4. Malli 4: Aivoinfarkti-, verenpaine- ja sydänsairauskuolleisuus ilman interaktioita. Devianssista selitetty 33,6 %.

Parametriset termit			
Muuttujan nimi	Estimaatti	P-arvo	
Vakio	-6,54625	< 2e-16	***
Sukupuoli	1,08484	< 2e-16	***
Siviilisääty	0,59597	3,51e-11	***
Depressio	-0,32388	0,113	
Ahdistus	1,18050	5,95e-08	***
Alkoholi	-0,39819	0,174	
Tupakka	0,53753	2,47e-06	***
Liikunta	-0,80152	< 2e-16	***
Metabolinen	0,63989	4,42e-12	***
Tasoitettut termit			
Muuttujan nimi	Edf	P-arvo	
TAS-summa	7,583	3,08e-16	***
Ikä	8,636	< 2e-16	***
Koulutus	3,520	0,000942	***
Painoindeksi	8,296	2,20e-11	***

(***) 0,001 (***) 0,01 (*) 0,05 (.) 0,1.

Taulukko 5. Interaktiomalli 1: Yleinen kuolleisuus. Devianssista selitetty 30,9 %.

Parametriset termit			
Muuttujan nimi	Estimaatti	P-arvo	
Vakio	1,52384	0,612674	
Koulutus	-0,50025	0,052659	.
Sukupuoli	0,89385	< 2e-16	***
Siviilisääty	0,52173	8,1e-10	***
Depressio	0,37744	0,021660	*
Ahdistus	0,62769	0,001204	**
Alkoholi	0,36943	0,045885	*
Tupakka	0,87515	< 2e-16	***
Liikunta	-0,30377	0,000407	***
Metabolinen	0,34466	0,000173	***
Tasoitettut termit			
Muuttujan nimi	Edf	P-arvo	
TAS-summa	6,839	4,45e-05	***
Ikä	2,408	1,46e-06	***
Painoindeksi	7,229	0,02883	*
Interaktio:TAS-summa,ikä	11,995	0,00127	**
Interaktio:TAS-summa,koulutus	5,962	0,00182	**
Interaktio:TAS-summa,painoindeksi	7,415	0,02408	*

(***) 0,001 (**) 0,01 (*) 0,05 (.) 0,1.

Taulukko 6. Interaktiomalli 2: Sydän- ja verisuonitautikuolleisuus. Devianssista selitetty 37,7 %.

Parametriset termit			
Muuttujan nimi	Estimaatti	P-arvo	
Vakio	-18,91145	0,8386	
Koulutus	1,03452	0,8967	
Sukupuoli	1,10332	< 2e-16	***
Siviilisääty	0,48157	1,19e-07	***
Depressio	-0,55602	0,0106	*
Ahdistus	1,43958	1,05e-11	***
Alkoholi	-0,33392	0,2256	
Tupakka	0,72461	9,92e-11	***
Liikunta	-0,73092	< 2e-16	***
Metabolinen	0,36894	7,93e-05	***
Tasoitettut termit			
Muuttujan nimi	Edf	P-arvo	
TAS-summa	7,970	< 2e-16	***
Ikä	8,221	< 2e-16	***
Painoindeksi	8,998	0,000254	***
Interaktio:TAS-summa,ikä	13,861	1,14e-07	***
Interaktio:TAS-summa,koulutus	17,635	6,23e-11	***
Interaktio:TAS-summa,painoindeksi	15,386	3,64e-12	***

(***) 0,001 (**) 0,01 (*) 0,05 (.) 0,1.

Taulukko 7. Interaktiomalli 3: Verenpaine- ja sydänsairauskuolleisuus. Devianssista selitetty 38,8 %.

Parametriset termit			
Muuttujan nimi	Estimaatti	P-arvo	
Vakio	-34,69000	< 2e-16	***
Sukupuoli	1,16351	< 2e-16	***
Siviilisääty	0,48325	2,34e-07	***
Depressio	-1,05605	1,43e-05	***
Ahdistus	1,49952	2,46e-11	***
Alkoholi	-0,30769	0,275	
Tupakka	0,50712	1,57e-05	***
Liikunta	-0,88244	< 2e-16	***
Metabolinen	0,61511	1,30e-10	***
Tasoitettut termit			
Muuttujan nimi	Edf	P-arvo	
TAS-summa	8,287	1,21e-10	***
Ikä	9,000	3,89e-14	***
Koulutus	8,883	5,89e-09	***
Painoindeksi	8,980	< 2e-16	***
Interaktio:TAS-summa,ikä	15,398	2,82e-13	***
Interaktio:TAS-summa,koulutus	15,885	< 2e-16	***
Interaktio:TAS-summa,painoindeksi	13,010	4,11e-10	***

(***) 0,001 (**) 0,01 (*) 0,05 (.) 0,1.

Taulukko 8. Interaktiomalli 4: Aivoinfarkti-, verenpaine- ja sydänsairauskuolleisuus. Devianssista selitetty 39,4 %.

Parametriset termit			
Muuttujan nimi	Estimaatti	P-arvo	
Vakio	-27,13883	< 2e-16	***
Sukupuoli	1,09866	< 2e-16	***
Siviilisääty	0,62293	1,90e-11	***
Depressio	-0,48134	0,02755	*
Ahdistus	1,29155	1,13e-08	***
Alkoholi	-0,55401	0,06479	.
Tupakka	0,38929	0,00109	**
Liikunta	-0,94119	< 2e-16	***
Metabolinen	0,62689	2,58e-11	***
Tasoitettut termit			
Muuttujan nimi	Edf	P-arvo	
TAS-summa	9,000	< 2e-16	***
Ikä	8,901	3,24e-13	***
Koulutus	8,921	2,55e-14	***
Painoindeksi	8,935	4,69e-07	***
Interaktio:TAS-summa,ikä	15,875	1,01e-14	***
Interaktio:TAS-summa,koulutus	15,560	< 2e-16	***
Interaktio:TAS-summa,painoindeksi	15,976	1,63e-12	***

(***) 0,001 (**) 0,01 (*) 0,05 (.) 0,1.

Liite C: Mallien vertailua

Malli	Devianssi	Df	Dev-erotus	Df-erotus	P-arvo
Yleinen kuolleisuus					
Alkuperäinen malli	1812,1	4978,9			
Interaktio:ikä	1795,7	4966,8	16,400	12,158	1,00e-05
Interaktio:koulutus	1792,3	4966,9	19,780	12,033	2,01e-07
Interaktio:painoindeksi	1797,2	4962,5	14,907	16,423	0,000658
Interaktio:kaikki	1774,3	4954,2	37,834	24,781	5,87e-12
Sydän- ja verisuonitautikuolleisuus					
Alkuperäinen malli	911,06	4970,0			
Interaktio:ikä	888,01	4958,0	23,050	12,016	<2,2e-16
Interaktio:koulutus	893,72	4958,4	17,344	11,592	1,94e-15
Interaktio:painoindeksi	886,69	4953,6	24,370	16,430	<2,2e-16
Interaktio:kaikki	847,66	4923,9	63,397	46,112	<2,2e-16
Verenpaine- ja sydänsairauskuolleisuus					
Alkuperäinen malli	627,22	4967,4			
Interaktio:ikä	611,78	4948,1	15,441	19,327	<2,2e-16
Interaktio:koulutus	599,33	4945,8	27,897	21,597	<2,2e-16
Interaktio:painoindeksi	614,57	4955,0	12,649	12,345	3,34e-16
Interaktio:kaikki	566,82	4917,6	60,399	49,837	<2,2e-16
Aivoinfarkti-, verenpaine- ja sydänsairauskuolleisuus					
Alkuperäinen malli	702,68	4969,0			
Interaktio:ikä	694,01	4955,6	8,663	13,396	3,39e-08
Interaktio:koulutus	680,61	4953,8	22,062	15,171	<2,2e-16
Interaktio:painoindeksi	699,12	4961,0	3,557	7,950	0,001366
Interaktio:kaikki	641,35	4913,8	61,326	55,133	<2,2e-16