Bayes-malli tyypin 2 diabetekseen liittyville tilasiirtymille

Tilastotieteen pro gradu -tutkielma

2. maaliskuuta 2020
Tiia-Maria Pasanen
Matematiikan ja tilastotieteen laitos
Jyväskylän yliopisto

JYVÄSKYLÄN YLIOPISTO
Matematiikan ja tilastotieteen laitos
Pasanen, Tiia-Maria: *Bayes-malli tyypin 2 diabetekseen liittyville tilasiirtymille*
Tilastotieteen pro gradu -tutkielma, 34 sivua, 4 liitettä (13 sivua)
2. maaliskuuta 2020

Tiivistelmä

Tyypin 2 diabetes on Suomessa kansantaudiksi luokiteltava glukoosiaineenvaihdunnan häiriö, joka aiheuttaa yhteiskunnalle jatkuvasti kasvavia kustannuksia ja heikentää sairastuneen elämänlaatua. Tauti todetaan tyypillisesti vasta vuosien sairastamisen jälkeen, jolloin erilaisia oheissairauksia on jo ehtinyt syntyä. Tyypin 2 diabetes esiintyy usein metabolisesta oireyhtymästä kärsivillä, ja sairastumiseen on mahdollista toisinaan vaikuttaa omilla elämäntavoilla. Esivaiheen diabetekseen eli prediabetekseen ja piilevänä sairastettavaan tyypin 2 diabetekseen sairastumisen riskin arvioimista asianmukaisten taustatekijöiden perusteella voitaisiin hyödyntää sekä yksilön että yhteiskunnan hyväksi.

Tässä tutkielmassa sovitetaan suomalaisia, vuosina 1884−1949 syntyneitä entisiä huippu-urheilijamiehiä ja heidän verrokkejaan kuvaavaan aineistoon tyypin 2 diabeteksen etenemistä kuvaava tilasiirtymämalli. Tämän mallin avulla estimoidaan siirtymätodennäköisyyksiä sekä tiettyjen taustamuuttujien vaikutusta niihin. Tilasiirtymämallin ajatuksena nimensä mukaisesti on, että on vaihtoehtoisia tiloja, joiden välillä voidaan siirtyä tietyn ajan kuluessa. Analysoitavan aineiston tulee sisältää tieto tilasta, jossa ollaan kullakin hetkellä. Tämän tutkielman malli on hierarkkinen, ja se sovitetaan käyttäen R- ja JAGS-ohjelmistoja. Tilasiirtymätodennäköisyyksiä ja taustamuuttujien vaikutusta estimoidaan multinomiaalisella logistisella regressiolla.

Sovitettava malli on yksinkertaistus laajemmasta tilasiirtymämallista, joka ottaa huomioon piilevän tyypin 2 diabeteksen tai prediabeteksen. Vaikka laajemman mallin teoriaa esitellään, sovitetaan aineistoon yksinkertaisempi malli, jonka avulla siirtymätodennäköisyyksien ja kovariaattien vaikutuksen estimointi näyttää onnistuvan hyvin. Tilasiirtymätodennäköisyyksien lisäksi aineistosta estimoidaan prediabeteksen tai piilevän tyypin 2 diabeteksen sairastamisen todennäköisyyttä vuonna 2008.

Avainsanat: Bayes-tilastotiede, huippu-urheilija, multinomiaalinen logistinen regressio, prediabetes, tilasiirtymä, tyypin 2 diabetes

Sisältö

1 Johdanto

Diabetes on koko ajan yleistyvä ryhmä eri syistä johtuvia ja eri tavalla ilmeneviä sairauksia, joiden yhteisenä oireena on veren glukoosipitoisuuden pitkäaikainen kohoaminen. Perinteisesti diabetes on jaettu kahteen eri alatyyppiin, tyypin 1 eli nuoruusiän diabetekseen ja tyypin 2 eli aikuisiän diabetekseen. Nykytiedon valossa tyyppejä on useita muitakin. (Tyypin 2 diabetes. Käypä hoito -suositus, 2018). Tässä tutkielmassa keskitytään perinteisen jaon mukaiseen tyypin 2 diabetekseen, johon sairastumiseen voidaan ainakin osittain vaikuttaa omilla elämäntavoilla.

Elämäntapojen vaikutukseen on yhteydessä erityisesti metabolinen oireyhtymä eli terveydentila, johon voivat liittyä kohoamat glukoosin paastoarvossa, veren triglyseridipitoisuudessa ja verenpainearvoissa sekä lasku veren hyvän eli HDL-kolesterolin pitoisuudessa. Metabolinen oireyhtymä voidaan todeta, jos henkilöllä on keskivartalolihavuutta ja ainakin kaksi edellä mainituista poikkeamista. (Alberti ym., 2006). Tyypin 2 diabetes ilmenee tyypillisesti yhdessä metabolisen oireyhtymän kanssa, ja liikunnan merkitys tyypin 2 diabeteksen kehittymiselle ja ilmenemismuodolle on siten merkittävä. (Roberts ym., 2013). Tämän tutkielman tarkoituksena on sovittaa analysoitavaan aineistoon Bayes-tilasiirtymämalli, joka kuvaa tyypin 2 diabetekseen sairastumista ja taudin etenemistä, ja tutkia liikunnan ja muiden taustatekijöiden vaikutusta sairauden etenemisen todennäköisyyteen. Tarkastelun kohteena on aineisto vuosina $1884-1949$ syntyneistä suomalaisista miehistä, joista noin puolella on huippu-urheilijatausta (Sarna ym., 1993). Aineistoa esitellään tarkemmin seuraavassa luvussa.

Aikuisiän diabetes voi johtua kahdesta eri insuliinin tuotantoon ja toimintaan liittyvästä häiriöstä. Kun elimistön oma insuliinituotanto heikentyy eikä riitä enää kattamaan koko insuliinin tarvetta, voi tila lopulta johtaa sairastumiseen. Toinen mahdollinen syy taudin syntyyn on elimistön heikentynyt insuliiniherkkyys eli kudosten alentunut kyky hyödyntää elimistössä jo olevaa insuliinia. Myös insuliinin tuotannon ja elimistön insuliiniherkkyyden yhtäaikainen yhdistelmä voi aiheuttaa

sairastumisen. Taudin puhkeamiseen liittyy geneettistä alttiutta, mutta siihen on joissain tilanteissa mahdollista vaikuttaa myös omilla elämäntavoilla, kuten ylipainon välttämisellä ja terveellisellä ruokavaliolla. Veren glukoosipitoisuuden kohoamisen syistä erityisesti elimistön insuliiniherkkyyden heikentymiseen voi vaikuttaa omalla toiminnallaan. (Scobie ja Samaras, 2009, 24-29).

Koska tyypin 2 diabetes kehittyy yleensä hitaasti, se havaitaan usein vasta vuosien sairastamisen jälkeen. Tällöin diagnoosin saamisen aikaan sairastuneella esiintyy jo taudin aiheuttamia komplikaatioita, jotka heikentävät elämänlaatua ja joiden hoito aiheuttaa yhteiskunnalle kustannuksia. (Scobie ja Samaras, 2009; Tyypin 2 diabetes. Käypä hoito -suositus, 2018). Siten olisi hyödyllistä löytää keinoja sairastumisen ehkäisemiseen ja oireiden vähentämiseen.

Suomessa diabetes luokitellaan nykyään kansantaudiksi, jota sairastaa arviolta noin 400 000 ihmistä. Näistä 350 000 on nimenomaan tyypin 2 diabeetikkoja. Tämän lisäksi taudin epäillään olevan piilevänä noin 100 000 henkilöllä. Lääkekorvauksia saavien määrä kasvaa jatkuvasti eli kyse on laajenevasta ongelmasta. (Diabetesliitto, 2019).

Liikunta ehkäisee metabolista oireyhtymää, ja sen on huomattu parantavan solujen insuliiniherkkyyttä (Roberts ym., 2013). Siksi on mielenkiintoista tutkia, kuinka tutkittavan aineiston sisältämä tieto henkilön huippu-urheilijataustasta ja siten poikkeuksellisen liikunnallisesta menneisyydestä vaikuttaa vielä vanhemmalla iällä diabetekseen sairastumiseen. Myös liikunnan määrä tutkimusajankohtana huomioidaan vapaa-aikana harrastetun liikunnan voimakkuutena.

Tässä tutkielmassa diabeteksen etenemistä mallinnetaan tilasiirtymämallin avulla. Tämä tarkoittaa, että henkilön terveydentila tyypin 2 diabeteksen suhteen jaetaan eri luokkiin, henkilö voi olla esimerkiksi terve tai diabeetikko. Näitä luokkia kutsutaan tiloiksi, joiden välillä henkilöt voivat liikkua ajan myötä. Kaikkia muutoksia tiloista toiseen tai samassa tilassa pysymistä kutsutaan siirtymiksi. Tilasiirtymiä on hyödynnetty myös esimerkiksi diabeteksen yleisyyttä Yhdysvalloissa vuodelle 2050 ennustavassa tutkimuksessa (Honeycutt ym., 2003), jossa mahdollisia tiloja ovat olleet terve, diabeetikko ja kuollut.

Tilasiirtymämallin siirtymätodennäköisyyksien estimoimiseksi aineistoon sovitetaan hierarkkinen Bayes-malli, jonka osana ovat kovariaattien vaikutusta estimoivat multinomiaaliset regressiomallit. Multinomiaalista logistista regressiomallinnusta on sovellettu diabetestutkimukseen aikaisemminkin. Esimerkiksi vuonna 2016 julkaistussa tutkimuksessa on tutkittu terveiden, prediabeetikkojen ja diabeetikkojen osuuksia multinomiaalisen logistisen regression avulla floridalaisessa aikuisväestössä (Okwechime ym., 2016). Samaan tapaan on estimoitu tyypin 2 diabeteksen esiintyvyyttä Tianjinissa asuvilla 20-79-vuotiailla kiinalaisilla (Xu ym., 2012). Koska tässä tutkielmassa halutaan tietoa todennäköisyyksistä, joilla henkilö siirtyy tilasta toiseen, eikä pelkästään eri tiloihin kuuluvien osuuksista, käytetään multinomiaalisia regressiomalleja koko rakennettavan mallin osana.

Sekä Floridaa että Kiinaa koskevissa tutkimuksissa normaalia korkeamman painoindeksin on havaittu liittyvän sekä diabeteksen että prediabeteksen kohonneeseen esiintymistodennäköisyyteen. Myös vähäisen liikunnan määrän on havaittu molemmissa tutkimuksissa olevan yhteydessä kohonneeseen diabetekseen sairastumisen riskiin. (Okwechime ym., 2016; Xu ym., 2012). Floridassa vähäisellä liikunnan määrällä on havaittu olevan yhteys kasvaneeseen prediabeteksen riskiin, mutta toisaalta tutkimuksen mukaan yhteys saattaa olla sattumaa (Okwechime ym., 2016). Yhteyden selittyminen pelkällä sattumalla vaikuttaa kuitenkin ristiriitaiselta ajatellen painoindeksin ja liikunnan määrän keskinäistä yhteyttä. Iällä on näyttänyt olevan kasvattava vaikutus Kiinassa sekä tyypin 2 diabeteksen että prediabeteksen havaitsemisen riskiin (Xu ym., 2012) ja tyyppien 1 tai 2 diabeteksen sairastamisen riskiin Floridassa (Okwechime ym., 2016). Iän ja painoindeksin vaikutusta taustamuuttujana on tarkoitus tarkastella myös huippu-urheilijaaineistosta.

Seuraavassa luvussa esitellään tutkimuskysymystä ja tutkittavaa aineistoa. Luvussa 3 käydään läpi mallinnukseen käytettäviä menetelmiä, mahdollisia tiloja ja siirtymiä, sekä vuosittaisten siirtymätodennäköisyyksien laskemista. Luvussa 4 kerrotaan saaduista tuloksista ja viimeisessä luvussa tehdään yhteenveto ja pohditaan tutkimuksen heikkouksia ja vahvuuksia.

2 Tutkimuskysymys ja tutkittava aineisto

Seuraavaksi esitellään tutkimusongelma ja kuvaillaan tässä tutkielmassa käytettävää huippu-urheilija-aineistoa ja sen keräämistä. Sen lisäksi esitellään analyysin kohteeksi valittuja muuttujia ja niiden käsittelyä tässä tutkielmassa.

2.1 Tutkimuskysymys

Tämän työn tarkoituksena on tutkia seuraavaksi esiteltävästä aineistosta, millaisia siirtymätodennäköisyyksiä diabetekseen liittyvien eri terveydentilojen välillä on ja kuinka eri taustamuuttujat niihin vaikuttavat. Koska käytettävissä on aineisto, joka kuvaa muun muassa tyypin 2 diabetekseen sairastumisen vaiheita, on tavoitteena tutkia, kuinka todennäköisesti tiettyjen taustamuuttujien vaikutuksen huomioiden henkilölle alkaa kehittyä diabetes, hän sairastuu tai kuolee. Tavoitteena on rakentaa tilasiirtymämalli, joka kuvaa henkilön terveydentilan kehittymistä tutkimuksen aikana. Tilat kuvaavat kunakin tutkimusvuotena havaittuja henkilöiden statuksia diabeteksen suhteen. Vuosien kuluessa on mahdollista edetä tilasta toiseen sairastumisen tai sairauden edetessä tai henkilön kuollessa. Kaikki nämä tilojen muutokset tai samassa tilassa pysyminen vastaavat mallissa siirtymiä.

Aineisto keskittyy liikuntaan ja terveyteen, ja kuten aiemmin kerrottiin, tyypin 2 diabetekseen sairastumiseen on toisinaan mahdollista vaikuttaa omilla elintavoillaan, joten luonnollista on valita tutkittaviksi taustamuuttujiksi juuri liikunnallisesta aktiivisuudesta kertovia muuttujia, kuten mahdollinen entinen huippu-urheilijatausta. Tarkoituksena on pystyä kertomaan, kasvattavatko vai pienentävätkö valitut taustamuuttujat diabeteksen etenemisen riskiä tämän aineiston perusteella. Muuttujia, joiden vaikutusta tutkitaan, kuvaillaan tarkemmin omassa alaluvussaan myöhemmin.

2.2 Aineisto

Analysoitavana on huippu-urheilijoista ja heidän verrokeistaan koostuva pitkittäisaineisto, jota on kerätty Entisten kilpaurheilijoiden seurantatutkimus -hankkeessa (Finnish male former elite athlete cohort). Hanke on Helsingin ja Jyväskylän yliopistojen sekä Terveyden ja hyvinvoinnin laitoksen yhteistyötutkimus. Aineistoa on kerätty sekä kyselylomakkeilla (Kujala ym., 2003) että laboratoriomittauksilla (Kujala ym., 2016). Tutkimushenkilöiksi on alun perin valittu urheilulajeittain suomalaisia miesurheilijoita sekä heitä iältään ja kotikunnaltaan 20-vuotiaina vastanneita verrokkihenkilöitä, jotka eivät ole ammattimaisia kilpatason huippu-urheilijoita. Kaikki tutkittavat henkilöt ovat miehiä, jotka on luokiteltu terveiksi 20-vuotiaina. Mukaan valitut urheilijat ovat edustaneet Suomea olympialaisissa, maailman- tai euroopanmestaruuskisoissa tai muissa kahden tai kolmen eri maan välisissä urheilukilpailuissa vähintään kerran vuosina 1920-1965. Heidän lajeinaan ovat olleet yleisurheilu, hiihto, jalkapallo, jääkiekko, koripallo, nyrkkeily, paini, painonnosto ja ammunta. Alkuperäisten tutkimushenkilöiden ja verrokkien valinta on tehty vuosina 1978−1979. Alun perin tutkittaviksi valittujen henkilöiden lisäksi mukaan on otettu vielä tiettyjen urheilulajien edustajia, joille ei kuitenkaan ole valittu enää verrokkeja, mistä johtuvat urheilijoiden ja verrokkien eroavat määrät jo tutkimuksen alussa. Koko aineisto sisältää yhteensä 3532 tutkimushenkilöä, joista 2078 on entisiä huippu-urheilijoita ja 1453 verrokkeja. Lisäksi yhden henkilön urheilustatus on tuntematon. Tutkimushenkilöiden valitsemisesta ja aineiston keräämisestä kerrotaan tarkemmin Sarnan ym. artikkelissa (1993).

Aineistoa on kerätty kyselylomakkeilla vuosina 1985, 1995, 2001 ja 2008. Lomakkeilla on kartoitettu erilaisia terveydentilaan ja liikuntaan sekä henkilön omaan taustaan liittyviä muuttujia. Kaikkien kyselykertojen lomakkeet eroavat toisistaan, mutta tässä työssä keskitytään muuttujiin, joita on kysytty jokaisella kerralla. Lisäksi vuonna 2008 vielä elossa olleille, vähintään yhteen aikaisemmista kyselylomakkeista vastaneille henkilöille lähetettiin kyselylomakkeen lisäksi kutsu terveydentilaa kartoittaviin laboratoriomittauksiin. Kutsuttujen ja mittauksiin paikalle

saapuneiden henkilöiden laboratoriotuloksia on tässä aineistossa käytössä 597 henkilöltä. Laboratoriokokeiden tuloksista hyödynnetään veren glukoosipitoisuusmittauksia.

2.3 Tutkittavat muuttujat

Kiinnostavia muuttujia ovat tutkimushenkilön ikä, sosioekonominen status, urheilutausta sen mukaan, onko tutkimushenkilö entinen huippu-urheilija vai ei, painoindeksi, sekä diabeteksen tilaa kuvaava muuttuja. Liikunnan vaikutuksesta kertoo myös MET eli metabolinen ekvivalentti, joka kuvaa liikunnallisen aktiivisuuden raskaustasoa (Kujala ym., 1998).

Tilat, joiden välisistä siirtymistä tässä tutkielmassa halutaan saada tietoa, kuvaavat tutkimushenkilöiden terveydentilaa tyypin 2 diabeteksen suhteen. Näiden tilojen, diabetesstatusten, mahdollisia arvoja ovat terve, diabeetikko ja kuollut. Lisäksi on erillinen riskiryhmään kuuluvien ryhmä, jonka muodostamisesta kerrotaan myöhemmin enemmän. Diabetesstatus on muodostettu useamman eri muuttujan avulla. Vuosilta 1985, 1995 ja 2001 käytetään ainoastaan kyselylomakkeella kerättyä tietoa siitä, onko lääkäri todennut tutkimushenkilöllä koskaan diabetesta. Mikäli ei ole, henkilö luokitellaan terveeksi tyypin 2 diabeteksen suhteen. Vastauksissa on eritelty, minkä tyypin diabeteksesta on ollut kyse, ja tämän aineiston tutkittavat henkilöt voivat sairastaa vain tyypin 2 diabetesta. Mikäli henkilö on vastannut jonakin vuonna saaneensa lääkäriltä tyypin 2 diabetesdiagnoosin, kaikkina myöhempinä kyselykertoina hänen oletetaan edelleen olevan diabeetikko, ellei hän kuole, jolloin tila merkitään sen mukaiseksi. Tämä yksinkertaistus vastaa oletusta, etteivät tyypin 2 diabeetikot voi parantua diagnoosin saamisen jälkeen.

Vuodelta 2008 hyödynnetään laboratoriomittauksin kerättyjä veren glukoosiarvoja. Laboratoriomittausten tulokset on jaettu alun perin viiteen luokkaan, jotka tässä yhdistetään kolmeksi kokoavaksi luokaksi. Uusista luokista ensimmäiseen kuuluvat henkilöt ovat diabeteksen kannalta terveitä eli heidän glukoositasonsa ovat normaaleja. Toiseen luokkaan kuuluvilla henkilöillä on aikaisemmin todettu

tyypin 2 diabetes. Kolmanteen luokkaan yhdistetään henkilöt, joilla voidaan havaita laboratoriossa tehdyillä mittauksilla tyypin 2 diabetes tai niin kutsuttu prediabetes (Saukkonen, 2012) heikentyneen glukoositoleranssin tai kohonneen glukoosipitoisuuden paastoarvon perusteella. Prediabetes voidaan todeta myös muilla perusteilla, mutta tässä huomioidaan vain glukoositoleranssin heikentyminen ja kohonnut paastoarvo. Prediabeteksessa poikkeamat glukoositoleranssissa ja paastoarvoissa eivät kuitenkaan ole niin suuria, että henkilön voitaisiin sanoa sairastavan diabetesta. Havainnot, joissa on erilaisia poikkeamia suhteessa terveen henkilön glukoosiarvoihin, yhdistetään yhteen riskiryhmäluokkaan, koska tällöin muodostuu ryhmä, jonka edustajat sairastavat tietämättään tyypin 2 diabetesta tai heillä on viitteitä siitä. Riskiryhmään kuuluvat henkilöt tiedetään vain vuodelta 2008. Jos tutkittavan glukoosiarvot on todettu normaaleiksi vuonna 2008, oletetaan hänen olleen terve aikaisempinakin vuosina. Tämän tutkielman analyyseissa riskiryhmään kuulumista käsitellään latenttina muuttujana, ja riskiryhmän henkilöt sisällytetään terveiden ryhmään. Mikäli henkilöltä tunnetaan lomakevastaus vuoden 2008 diabetesstatuksesta, mutta ei laboratoriomittauksen tulosta, hyödynnetään tällöin ainoastaan lomakkeella kerättyä tietoa. Kokonaisuudessaan lomakkeiden ja laboratoriomittausten avulla muodostetun diabetesstatusmuuttujan eri luokkiin kuuluvien määriä on havainnollistettu taulukossa 1.

Taulukko 1: *Eri diabetesstatusluokkiin kuuluvien määrät vuosittain. Sarakkeessa Normaali on terveeksi luokiteltujen ja sarakkeessa Diagnoosi tyypin 2 diabeteksen diagnoosin saaneiden lukumäärä. Vuoden 2008 terveiden lukumäärä koostuu terveiksi luokiteltujen sekä riskiryhmään luokiteltujen määristä, jotka on esitetty suluissa vastaavassa järjestyksessä. Sarakkeessa Kuollut on kuolleiden kumulatiivinen lukumäärä ja sarakkeessa Puuttuu diabetesstatukseltaan tuntemattomattomien henkilöiden lukumäärä.*

Vuosi	Normaali	Diagnoosi	Kuollut	Puuttuu
1985	1935	81	701	716
1995	1404	93	1336	600
2001	1100	103	1671	559
2008	466 (215, 251)	178	2064	725

Analyyseihin käytettävä aineisto sisältää seuraavaksi kuvailtavat muuttujat henkilöiltä, joiden diabetesstatus on terve tai diabeetikko kunkin tutkimusvuoden lähtötilanteessa. Lopulta jää siis analysoitavaksi 2016 siirtymää eli havaintoa vuodelta 1985, 1497 havaintoa vuodelta 1995 ja 1203 havaintoa vuodelta 2001. Tilaa, johon henkilö on seuraavaan tutkimuskertaan mennessä siirtynyt, ei välttämättä tiedetä. Tämä tarkoittaa, että seuraavissa muuttujia kuvaavissa taulukoissa esitetyt lukumäärät ja tunnusluvut on laskettu niiden henkilöiden perusteella, jotka ovat kunkin siirtymänsä lähtövuosina 1985, 1995 tai 2001 terveitä tai diabeetikkoja (ks. taulukko 1).

Yhteensä 99 tutkimushenkilön ikää ei tunneta, ja heidät jätetään tämän analyysin ulkopuolelle muutenkin puutteellisten vastausten ja taustatietojen vuoksi. Näihin pois jääviin henkilöihin lukeutuu myös urheilustatukseltaan tuntematon. Mikäli henkilön ikä puuttuu vain joltain tutkimusvuodelta, täydennetään se muuna kyselykertana havaitulla iällä, johon lisätään tai vähennetään tutkimusvuosien ajallinen ero kokonaisina vuosina.

Sosioekonominen status on kysytty vuonna 1985, minkä jälkeen tietoa ei ole päivitetty eli se on määräytynyt ensimmäisen vastauskerran mukaisesti. Alun perin tiedusteltu sosioekonominen status on voinut olla johtaja, toimistotyöntekijä, ruumiillisen työn tekijä, kouluttautumaton, maanviljelijä tai muu. Tässä luokkien määrää on vähennetty kolmeen, joista yhteen lukeutuvat johtajat, toiseen toimistotyöntekijät ja kolmanteen kaikki loput. Eri sosioekonomisiin luokkiin kuuluvien määrät ovat taulukossa 2.

Taulukko 2: *Eri sosioekonomisiin luokkiin kuuluvien määrät vuosittain. Puuttuu-sarakkeessa puuttuvien havaintojen lukumäärä.*

Vuosi	Muu	Toimistotyöntekijä	Johtaja	Puuttuu
1985	900	711	401	4
1995	629	516	327	25
2001	479	435	278	11
2008	245	234	159	6

Tutkimushenkilön liikuntahistoriasta kertoo jo tutkittavia valittaessa kartoitettu ja valinnan kriteerinäkin ollut urheilustatus. Urheilijoiden ja verrokkien lukumäärät on taulukoitu taulukkoon 3. Liikunnan määrästä tutkimuksen aikana on puolestaan kerätty tietoa MET-indeksin avulla. MET on indeksi, joka kuvaa tunnin mittaisen fyysisen aktiivisuuden rasittavuutta suhteessa istumiseksi määriteltyyn lepotilaan, jonka arvoksi on asetettu 1. Kävelemisen MET-arvo on noin 4, hölkkäämisen 10 ja juoksemisen 13, mikä tarkoittaa sitä, että esimerkiksi juokseminen tunnin ajan kasvattaa MET-arvoa 13 yksikköä. MET-arvosta ja sen laskemisesta kerrotaan Kujalan ym. artikkelissa (1998). Tässä aineistossa laskettu MET-arvo kuvaa vapaa-aikana harrastetun liikunnan voimakkuutta viikossa. Aineiston pienimmät, alle lepotila-aktiivisuuden olevat MET-arvot ovat henkilöiltä, jotka raportoivat liikkuvansa hyvin vähän vapaa-ajallaan esimerkiksi liikuntakyvyttömyyden takia ja suurimmat henkilöiltä, jotka voivat harrastaa vielä eläkkeelläkin kilpaurheilua. Myös painoindeksin avulla pyritään huomioimaan terveellisten elämäntapojen vaikutusta analyysissa. MET-arvoon, painoindeksiin ja ikään liittyviä muttujien tietoja on kirjattu taulukkoon 4 erikseen urheilijoille ja verrokeille. Keskimäärin iät ja painoindeksit näyttävät olevan samaa kokoluokkaa urheilijoiden ja verrokkien kesken, mutta urheilijat näyttävät raportoivan keskimäärin liikkuvansa vapaa-ajallaan huomattavasti aktiivisemmin kuin verrokit.

Taulukko 3: *Urheilijoiden ja verrokkihenkilöiden lukumäärät analysoitavassa aineistossa eri tutkimusvuosina.*

Vuosi	Urheilija	Verrokki
1985	1251	765
1995	929	568
2001	780	423
2008	414	230

Taulukko 4: *Jatkuvien muuttujien minimit, keskiarvot ja maksimit sekä puuttuvien havaintojen määrät vuosittain urheilijoille ja verrokeille. MET-arvojen maksimi on useampana vuotena sama, ja tämän arvon saavuttavat henkilöt vaihtelevat vuosittain. Maksimin saavuttajia on joka vuosi enemmän kuin yksi. Myös saman minimin saavuttajat vaihtelevat vuosittain ja heitä on useampia jokaisena vuotena.*

Muuttuja	Selite	Minimi	Keskiarvo	Maksimi	Puuttuu
Urheilijat					
ika85	ikä vuonna 1985	35.90	56.70	93.70	0
BMI85	painoindeksi vuonna 1985	16.25	26.11	43.30	3
MET85	metabolinen ekvivalentti vuonna 1985	0.04	30.57	227.50	19
ika95	ikä vuonna 1995	46.13	64.12	94.51	0
BMI95	painoindeksi vuonna 1995	16.07	26.31	46.30	9
MET95	metabolinen ekvivalentti vuonna 1995	0.10	30.62	227.50	33
ika01	ikä vuonna 2001	51.89	68.21	99.29	0
BMI01	painoindeksi vuonna 2001	17.76	26.26	46.30	9
MET01	metabolinen ekvivalentti vuonna 2001	0.04	28.64	227.50	37
ika08	ikä vuonna 2008	59.17	72.81	92.27	0
BMI08	painoindeksi vuonna 2008	19.60	26.52	52.60	39
MET08	metabolinen ekvivalentti vuonna 2008	0.04	31.35	175.00	31
Verrokit					
ika85	ikä vuonna 1985	38.10	54.92	86.40	0
BMI85	painoindeksi vuonna 1985	15.79	26.40	58.13	7
MET85	metabolinen ekvivalentti vuonna 1985	0.04	14.59	227.50	10
ika95	ikä vuonna 1995	48.25	62.07	85.37	0
BMI95	painoindeksi vuonna 1995	16.18	26.83	42.61	12
MET95	metabolinen ekvivalentti vuonna 1995	0.10	16.74	146.25	23
ika01	ikä vuonna 2001	54.03	66.86	93.28	0
BMI01	painoindeksi vuonna 2001	14.96	26.84	42.61	14
MET01	metabolinen ekvivalentti vuonna 2001	0.04	18.31	146.25	29
ika08	ikä vuonna 2008	61.32	71.49	96.84	0
BMI08	painoindeksi vuonna 2008	19.00	26.79	37.80	25
MET08	metabolinen ekvivalentti vuonna 2008	0.04	20.54	175.00	29

3 Menetelmät

Tässä luvussa kuvaillaan, kuinka mallinnettavia siirtymätodennäköisyyksiä ja taustamuuttujien vaikutuksia estimoidaan. Aluksi esitellään mahdolliset tilat, siirtymät ja niihin liittyvät merkinnät. Sen jälkeen käydään läpi siirtymätodennäköisyyksien laskemista, kun yhden vuoden aikana tapahtuvan siirtymän todennäköisyys tunnetaan. Kuvaillaan myös vuosittaisten todennäköisyyksien estimoinnin taustalla olevaa teoriaa ja taustamuuttujien vaikutuksen huomioimista multinomiaalisella logistisella regressiolla. Lopuksi esitellään mallit, jotka aineistoon sovitetaan.

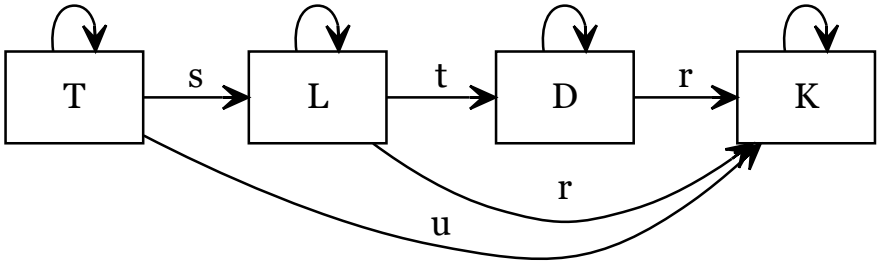
3.1 Tilat ja siirtymät

Tässä mahdolliset tilat perustuvat diabetesstatuksesta kertovaan muuttujaan. Vuosina 1985, 1995 ja 2001 havaitut tilat kertovat, onko henkilö terve, diagnosoitu tyypin 2 diabeetikko vai kuollut. Vuonna 2008 kerätty tieto diabetesstatuksesta lisää mahdollisten tilojen joukkoa riskiryhmään kuulumisen tilalla. Tutkittavassa aineistossa ei kuitenkaan ole tietoa siitä, kuuluuko henkilö vuosina 1985, 1995 tai 2001 terveisiin vai riskiryhmään. Mahdollisia tiloja on siis neljä.

Mallissa 1 ensimmäinen tila kuvaa diabeteksen suhteen tervettä henkilöä, jolla ei havaita edes viitteitä tyypin 2 diabeteksesta. Toinen tila on riskiryhmän tila, jossa henkilöllä on diagnosoimaton tyypin 2 diabetes tai kohonneiden glukoosiarvojen antamia viitteitä siitä. Kolmas tila edustaa diagnosoituja tyypin 2 diabeetikkoja ja neljäs kuolleita. Koska mallinnettavat tilat kuvaavat elämän ja mahdollisen sairastumisen kulkua ja koska yksinkertaisuuden vuoksi tässä oletetaan, ettei tyypin 2 diabeteksesta tai sen esivaiheesta ole mahdollista parantua, riippuvat mahdolliset siirtymät henkilön lähtötilasta.

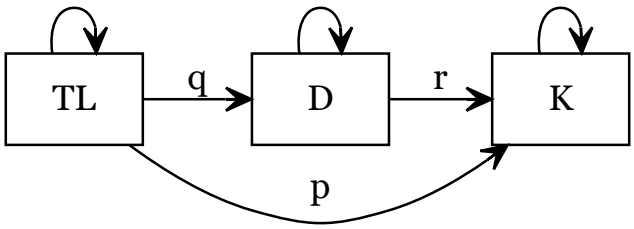
Kuvassa 1 on havainnollistettu mallin 1 mukaisia mahdollisia tiloja ja siirtymiä, joiden oletetaan kuvaavan etenevää prosessia, jolloin paluusiirtymät tiloihin,

joista on jo kerran poistuttu, ovat mahdottomia. Terve voi pysyä terveenä eli tilassa T, siirtyä riskiryhmään eli tilasta T tilaan L tai kuolla eli siirtyä tilasta T tilaan K. Riskiryhmän edustaja puolestaan voi pysyä riskiryhmässä eli tilassa L, siirtyä diabeetikoksi eli tilasta L tilaan D tai kuolla eli siirtyä tilasta L tilaan K. Diabeetikko voi pysyä diabeetikkona eli tilassa D tai kuolla, mikä vastaa siirtymää tilasta D tilaan K. Kuollut voi ainoastaan pysyä omassa tilassaan eli tilassa K. Riskiryhmään kuuluvan ja diagnosoidun tyypin 2 diabeetikon kuolemisen todennäköisyydet oletetaan samoiksi. Aikaa yhteen siirtymään tilasta toiseen oletetaan kuluvan vuosi.



Kuva 1: *Mallin 1 mukainen latentin muuttujan sisältävä tilasiirtymäkuvaaja. Kuvassa T vastaa terveitä, L riskiryhmään kuuluvia, D diabeetikkoja ja K kuolleita. Mahdollisia siirtymiä on kuvattu nuolilla, joiden viereen on merkitty vastaavia vuosittaisia siirtymätodennäköisyyksiä merkitsevät kirjaimet.*

Koska tutkittava aineisto ei sisällä tarpeeksi havaintoja piilevään tyypin 2 diabetekseen tai prediabetekseen eli riskiryhmään liittyvistä tiloista, täytyy mallia 1 yksinkertaistaa. Yksinkertaisemmassa tilasiirtymämallissa, jota kutsutaan jatkossa malliksi 2, riskiryhmän edustajille ei ole omaa tilaa, vaan heidät on sisällytetty terveiden ryhmään. Mallin 2 mukaisia mahdollisia tiloja ja siirtymiä on havainnollistettu kuvassa 2. Terve henkilö voi pysyä terveenä eli tilassa TL, sairastua eli siirtyä tilasta TL tilaan D tai kuolla eli siirtyä tilasta TL tilaan K. Tyypin 2 diabeetikko voi pysyä diabeetikkona eli tilassa D tai kuolla, mikä vastaa siirtymää tilasta D tilaan K. Kuolleet luonnollisesti pysyvät kuolleina eli tilassa K. Tässäkin oletetaan yhteen siirtymään kuluvan aikaa vuosi.



Kuva 2: *Mallin 2 mukainen tilasiirtymäkuvaaja ilman erillistä latenttia riskiryhmämuuttujaa. Kuvassa TL vastaa terveitä ja riskiryhmään kuuluvia, D diabeetikkoja ja K kuolleita. Nuolet kuvaavat mahdollisia siirtymiä ja niiden suuntia, ja niiden vierelle on merkitty kirjaimet, jotka kuvaavat vastaavia vuosittaisia siirtymätodennäköisyyksiä.*

Edellä on kuvailtu useita siirtymiä, joiden todennäköisyyksille käytetään omia merkintöjään. Kuvissa 1 ja 2 näkyvät nuolet edustavat kaikkia mahdollisia vuoden aikana tapahtuvia siirtymiä, ja kirjainmerkinnät siirtymänuolien vierellä tarkoittavat vuosittaisen siirtymän todennäköisyyttä. Useamman vuoden aikana on siis mahdollista käydä läpi kaikki tilat. Yhden vuoden siirtymätodennäköisyyksiä kuvataan kirjaimilla, joihin ei ole merkitty alaindeksiä, muuten alaindeksissä oleva luku kuvaa, kuinka monen vuoden aikana tapahtuvan siirtymän todennäköisyydestä on kyse. Esimerkiksi merkintä p tarkoittaa terveen henkilön todennäköisyyttä kuolla yhden vuoden aikana ja p_6 todennäköisyyttä, että terve henkilö kuolee kuuden vuoden kuluessa. Siirtymätodennäköisyyksien lisäksi on tarpeen kuvata omalla merkinnällään, l, todennäköisyyttä henkilön kuulumiselle riskiryhmään kunakin tutkimusvuotena. Vuodelta 2008 tämän muuttujan arvoista on havaintoja, mutta edellisinä tutkimuskertoina asiaa ei ole tutkittu. Riskiryhmään kuulumista käsitelläänkin latenttina muuttujana. Kaikki edellä mainitut todennäköisyysmerkinnät selitteineen on koottu taulukkoon 1.

Taulukko 5: *Tutkittavista siirtymätodennäköisyyksistä sekä riskiryhmään kuulumisen todennäköisyydestä käytettävät merkinnät selitteineen. Ilman alaindeksiä esiintyvät kirjaimet merkitsevät yhden vuoden siirtymätodennäköisyyttä. Muuten alaindeksissä oleva luku kuvaa, kuinka monen vuoden siirtymätodennäköisyydestä on kyse.*

Merkintä	Selitettävä todennäköisyys
p	terve tai riskiryhmän edustaja kuolee
q	terve tai riskiryhmän edustaja sairastuu tyypin 2 diabetekseen
r	tyypin 2 diabeetikko tai riskiryhmän edustaja kuolee
s	terve siirtyy riskiryhmään
t	riskiryhmän edustaja sairastuu tyypin 2 diabetekseen
u	terve kuolee
l	henkilö kuuluu riskiryhmään

3.2 Useaa vuotta koskevien todennäköisyyksien laskeminen

Edellä esitellyt siirtymätodennäköisyydet kuvaavat vuoden aikana tapahtuvan siirtymän todennäköisyyttä, mutta käytännössä tietoa tutkittavista on kerätty kymmenen, kuuden ja seitsemän vuoden välein. Eriävät vuosimäärät otetaan huomioon siirtymätodennäköisyyksiä laskettaessa.

Seuraavaksi käydään läpi mallin 2 mukaisten tilasiirtymätodennäköisyyksien laskemista. Oletetaan, että siirtymään on käytettävissä n kokonaista vuotta ja halutaan tietää todennäköisyys siirtymälle tilasta TL tilaan K. Kun p kuvaa terveen tai riskiryhmään kuuluvan vuosittaista todennäköisyyttä kuolla, q terveen tai riskiryhmään kuuluvan vuosittaista todennäköisyyttä sairastua tyypin 2 diabetekseen ja r diabeetikon vuosittaista todennäköisyyttä kuolla, lasketaan todennäköisyys, että terve henkilö on kuollut n vuoden kuluttua, kaavalla

$$p_n = \sum_{i=0}^{n-2} \sum_{j=0}^{n-2-i} (1-p-q)^j q (1-r)^i r + \sum_{k=0}^{n-1} (1-p-q)^k p.$$

Ensimmäinen summa vastaa siirtymäyhdistelmää, jossa henkilö pysyy terveenä tai riskiryhmässä j vuotta, sairastuu diabetekseen, sairastaa sitä i vuotta ja kuolee

sen jälkeen, mutta tietoa sairastumisesta ei koskaan kerätä. Kaavan jälkimmäinen summa kuvaa tilannetta, jossa henkilö pysyy terveenä tai riskiryhmässä k vuotta ja kuolee sitten.

Todennäköisyys siirtymälle tilasta TL tilaan D eli sille, että terve tai riskiryhmään kuuluva henkilö on n vuoden kuluttua sairastunut tyypin 2 diabetekseen saadaan puolestaan kaavasta

$$q_n = \sum_{i=0}^{n-1} (1-p-q)^i q (1-r)^{n-1-i},$$

missä n on siirtymään käytettävien kokonaisten vuosien määrä. Todennäköisyys lasketaan siis tilanteelle, jossa henkilö on terve tai riskiryhmässä i vuotta, sairastuu ja on sen jälkeen diabeetikko $n-1-i$ vuotta.

Todennäköisyys siirtymälle tilasta D tilaan K eli sille, että tyypin 2 diabeetikko kuolee, lasketaan kaavalla

$$r_n = \sum_{i=0}^{n-1} (1-r)^i r,$$

missä n on jälleen siirtymään käytettävissä olevien vuosien määrä. Henkilö siis pysyy diabeetikkona i vuotta, minkä jälkeen hän kuolee.

3.3 Multinomiaalinen logistinen regressio

Edellä olevissa kaavoissa tarvitaan yhden vuoden siirtymätodennäköisyyksiä, joita voidaan mallintaa multinomiaalisen logistisen regressiomallin avulla. Olkoon mahdollisten lähtötilojen joukko I ja lähtötilasta i riippuvien kohdetilojen joukko J_i. Merkitään henkilön tilaa tutkimuskerralla $t \in \{1,2,3,4\}$ kirjaimella Y_t. Olkoon lisäksi tilasta i tilaan j siirtymisen todennäköisyys tutkimuskerralla t havaittujen selittävien muuttujien arvojen, $\boldsymbol{x}_t$, ehdolla todennäköisyys

$$\pi_{ij}(\boldsymbol{x}_t) = \mathbb{P}(Y_{t+1} = j | Y_t = i,\, \boldsymbol{x}_t).$$

Kun referenssisiirtymäksi valitaan pysyminen lähtötilassa i, voidaan sen ja siirtymän tilaan j siirtymätodennäköisyyksien vetokertoimen logaritmia mallintaa regressiomallilla

$$\ln\left(\frac{\pi_{ij}(\boldsymbol{x}_t)}{\pi_{ii}(\boldsymbol{x}_t)}\right) = \boldsymbol{\beta}_{ij}^T \boldsymbol{x}_t. \tag{1}$$

Kertoimilla $\boldsymbol{\beta}_{ij}$ estimoidaan kovariaattien vaikutusta siirtymätodennäköisyyksiin yhden vuoden aikana. Soveltamalla eksponenttifunktiota ja ehtoa $\sum_{\hat{j}\in J_i} \pi_{i\hat{j}}(\boldsymbol{x}_t) = 1$ voidaan ratkaista yleisessä muodossa yksittäisen siirtymän todennäköisyys

$$\pi_{ij}(\boldsymbol{x}_t) = \frac{\exp(\boldsymbol{\beta}_{ij}^T \boldsymbol{x}_t)}{\sum_{\hat{j}\in J_i} \exp(\boldsymbol{\beta}_{i\hat{j}}^T \boldsymbol{x}_t)}.$$

Kun referenssisiirtymiä eli siirtymiä tilasta i tilaan i vastaavat regressiokertoimet asetetaan nolliksi, edellinen yhtälö sievenee muotoon

$$\pi_{ii}(\boldsymbol{x}_t) = \frac{1}{1 + \sum_{\hat{j}\in J_i \setminus i} \exp(\boldsymbol{\beta}_{i\hat{j}}^T \boldsymbol{x}_t)},$$

kun kyseessä ovat referenssisiirtymien todennäköisyydet, ja muotoon

$$\pi_{ij}(\boldsymbol{x}_t) = \frac{\exp(\boldsymbol{\beta}_{ij}^T \boldsymbol{x}_t)}{1 + \sum_{\hat{j}\in J_i \setminus i} \exp(\boldsymbol{\beta}_{i\hat{j}}^T \boldsymbol{x}_t)} \tag{2}$$

kaikkien muiden mahdollisten siirtymien todennäköisyyksille. Samaa teoriaa esimerkin avulla kuvaavat Hosmer ym. (2013, 270-271) ja yleisemmällä tasolla Dobson ja Barnett (2008, 149-151), jotka myös käyvät tarkemmin läpi estimoitavien todennäköisyyksien ja kategorisen jakauman yhteyttä toisiinsa.

Jotta siirtymätodennäköisyyksien laskemiseen tarvittavat regressiokertoimet saataisiin selville, sovitetaan aineistoon hierarkkinen Bayes-malli, joka huomioi vaihtelevat tutkimuskertojen väliset vuosimäärät. Olkoon $X_i = (\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \ldots, \boldsymbol{x}_{i_N})^T$ tilasta i tapahtuneita siirtymiä selittävien muuttujien matriisi, jossa N on tutkimushenkilöiden määrä, ja $Y = (y_{i1}, y_{i2}, \ldots, y_{iJ})$ tilasta i tilaan j havaittujen siir-

tymien lukumäärä. Tällöin

$$Y = (y_{i1}, y_{i2}, \ldots, y_{iJ}) \sim \text{Multinomi}(1, (\pi_{i1}, \pi_{i2}, \ldots, \pi_{iJ}))$$

eli yhden henkilön havaittujen siirtymien lukumäärät noudattavat multinomijakaumaa. Käytännössä se merkitsee tässä sitä, että lähtötilaltaan terveen henkilön siirtymien lukumäärät noudattavat kategorista jakaumaa,

$$Y_{TL} = (y_{\text{TLTL}}, y_{\text{TLD}}, y_{\text{TLK}}) \sim \text{Kategorinen}(1 - p_n - q_n, q_n, p_n),$$

missä n kuvaa taas kuluneiden vuosien lukumäärää. Alaindeksien kirjaimet ovat kuten kuvassa 2 eli TL vastaa terveitä, D diabeetikkoja ja K kuolleita. Lähtötilaltaan diabetesta sairastavan siirtymät puolestaan noudattavat bernoullijakaumaa parametrilla r_n eli

$$Y_D = (y_{\text{DD}}, y_{\text{DK}}) \sim \text{Bernoulli}(r_n),$$

missä n ja tilojen kirjainlyhenteet ovat kuten edellä.

3.3.1 Riskiryhmään kuulumisen todennäköisyys

Tutkitaan todennäköisyyttä, että henkilö kuuluu riskiryhmään vuonna 2008, kun selittävinä muuttujina käytetään vuonna 2001 havaittuja taustamuuttujien arvoja. Selittäviksi muuttujiksi valitaan edellisessä luvussa esitellyt muuttujat ikä (ika), painoindeksi (BMI), MET (MET), urheilijastatus (urh) ja sosioekonominen luokka. Sosioekonominen luokka jakautuu kahdeksi dikotomiseksi muuttujaksi, joista ensimmäinen kuvaa, onko henkilö toimistotyöntekijä vai ei ($toim$) ja toinen, onko henkilö johtaja vai ei ($joht$).

Oletetaan riskiryhmään kuulumisen noudattavan jakaumaa Bernoulli(l) ja riippuvan valittujen taustamuuttujien arvoista. Riskiryhmään kuulumisen todennä-

köisyyttä arvioidaan sovittamalla ensin kaavan 1 mukaisesti malli

$$\ln(\tilde{l}) = \boldsymbol{\beta}_{l_0} + \boldsymbol{\beta}_{l_{ika}} \cdot ika + \boldsymbol{\beta}_{l_{BMI}} \cdot BMI + \boldsymbol{\beta}_{l_{MET}} \cdot MET + \boldsymbol{\beta}_{l_{urh}} \cdot urh + \boldsymbol{\beta}_{l_{toim}} \cdot toim + \boldsymbol{\beta}_{l_{joht}} \cdot joht,$$

ja laskemalla sen jälkeen varsinainen todennäköisyys l kaavalla

$$l = \frac{\tilde{l}}{1 + \tilde{l}}.$$

Merkinnällä $\tilde{l}$ tarkoitetaan riskiryhmään kuulumisen todennäköisyyteen liittyvää vetokerrointa.

3.3.2 Malli 2: latentin riskiryhmätilan huomioimaton malli

Siirtymätodennäköisyyksien selvittämiseksi aineistoon sovitetaan kuvan 2 tilannetta vastaava malli 2, jossa riskiryhmään kuuluvat sisällytetään terveiden joukkoon. Kovariaateiksi valitaan siirtymän lähtötilassa havaitut muuttujat ikä (ika), painoindeksi (BMI), MET (MET), urheilijastatus (urh) ja sosioekonominen luokka, joka jakautuu dikotomisiksi muuttujiksi kuten edellä riskiryhmään kuulumisen todennäköisyyttä mallinnettaessa. Kaavan 1 mukaisesti sovitettavia regressiomalleja on tällöin kolme. Kaikki mallit kuvaavat taustamuuttujien vaikutusta yhteen vuoteen liittyviin siirtymätodennäköisyyksiin.

Taustamuuttujien vaikutusta terveen henkilön kuolemisen todennäköisyyteen estimoidaan yhtälöllä

$$\ln(\tilde{p}) = \boldsymbol{\beta}_{p_0} + \boldsymbol{\beta}_{p_{ika}} \cdot ika + \boldsymbol{\beta}_{p_{BMI}} \cdot BMI + \boldsymbol{\beta}_{p_{MET}} \cdot MET + \boldsymbol{\beta}_{p_{urh}} \cdot urh + \boldsymbol{\beta}_{p_{toim}} \cdot toim + \boldsymbol{\beta}_{p_{joht}} \cdot joht,$$

missä $\tilde{p}$ tarkoittaa terveen henkilön kuolemisen todennäköisyyteen liittyvää vetokerrointa.

Vastaavasti yhtälöllä

$$\ln(\tilde{q}) = \boldsymbol{\beta}_{q_0} + \boldsymbol{\beta}_{q_{ika}} \cdot ika + \boldsymbol{\beta}_{q_{BMI}} \cdot BMI + \boldsymbol{\beta}_{q_{MET}} \cdot MET + \boldsymbol{\beta}_{q_{urh}} \cdot urh + \boldsymbol{\beta}_{q_{toim}} \cdot toim + \boldsymbol{\beta}_{q_{joht}} \cdot joht,$$

missä $\tilde{q}$ merkitsee terveen sairastumisen todennäköisyyteen liittyvää vetokerrointa, estimoidaan taustatekijöiden vaikutusta terveen henkilön tyypin 2 diabetekseen sairastumisen todennäköisyyteen.

Kolmas yhtälö

$$\ln(\tilde{r}) = \boldsymbol{\beta}_{r_0} + \boldsymbol{\beta}_{r_{ika}} \cdot ika + \boldsymbol{\beta}_{r_{BMI}} \cdot BMI + \boldsymbol{\beta}_{r_{MET}} \cdot MET + \boldsymbol{\beta}_{r_{urh}} \cdot urh + \boldsymbol{\beta}_{r_{toim}} \cdot toim + \boldsymbol{\beta}_{r_{joht}} \cdot joht,$$

missä $\tilde{r}$ tarkoittaa diabeetikon kuolemisen todennäköisyyteen liittyvää vetokerrointa, puolestaan estimoi taustatekijöiden yhteyttä tyypin 2 diabeetikon kuolemisen todennäköisyyteen.

Varsinaiset siirtymätodennäköisyydet voidaan selvittää yllä laskettujen vetokerrointen ja kaavan 2 avulla käyttäen yhtälöitä

$$p = \frac{\tilde{p}}{1 + \tilde{p} + \tilde{q}}$$

$$q = \frac{\tilde{q}}{1 + \tilde{p} + \tilde{q}}$$

ja

$$r = \frac{\tilde{r}}{1 + \tilde{r}}.$$

Näin saatavien todennäköisyyksien kirjainmerkinnät vastaavat kuvaan 2 merkittyjä siirtymätodennäköisyyksiä.

4 Tulokset

Seuraavaksi esitellään mallinnuksesta saadut tulokset. Tavoitteena on siis estimoida diabeteksen etenemiseen liittyviä siirtymätodennäköisyyksiä ja taustatekijöiden vaikutusta niihin mallin 2 tapauksessa. Käydään aluksi läpi, millaisia estimaatteja riskiryhmään vuonna 2008 kuulumisen todennäköisyyteen l liittyville regressiokertoimille saadaan vuoden 2001 taustatietojen perusteella. Sen jälkeen kuvaillaan mallin 2 sovittamalla saatuja estimaatteja taustamuuttujien vaikutuksista ja siirtymätodennäköisyyksistä.

4.1 Riskiryhmään kuulumisen todennäköisyys

Tutkitaan, millä todennäköisyydellä luokan TL henkilö kuuluu riskiryhmään vuonna 2008, kun käytettävissä ovat kovariaattien arvot vuodelta 2001. Sovitettava luvun 3.3.1 mukainen malli on siis yksittäinen logistinen regressiomalli. Tutkittavista taustamuuttujista ikä, painoindeksi ja MET on keskistetty pelkästään vuoden 2001 havaintojen tasolla. Kovariaateissa on puuttuvaa tietoa, jonka imputoimiseen käytetään jakaumia, joita muuttujat näyttävät silmämääräisesti noudattavan. Aineistosta on arvioitu, että muuttujat noudattavat suurin piirtein jakaumia

$$BMI \sim N(0, 10),$$

$$MET \sim N(0, 700),$$

$$toim \sim \text{Bernoulli}(0.35),$$

ja

$$joht \sim \text{Bernoulli}(0.25),$$

kun jatkuvien muuttujien kohdalla on tehty keskistys ja normaalijakauman hajontaparametrina on varianssi. Muuttujat urh, $toim$ ja $joht$ saavat arvon 1, kun hen-

kilö kuuluu kyseiseen luokkaan, ja 0 muuten. Kaikkien regressiokertoimien prioriksi asetetaan normaalijakauma $N(0, 2)$, missä hajontaparametrina on varianssi.

Mallien sovittamiseen käytetään R-ohjelmistoa (R Core Team, 2019), JAGS-ohjelmaa (Plummer, 2003) ja niiden yhteistyön mahdollistavaa R-ohjelmiston lisäosaa, rjags-kirjastoa (Plummer, 2019). Mallin tuloksien tulkintaan käytetään myös R-ohjelmiston coda-kirjastoa (Plummer ym., 2006). Regressiokertoimille estimoidaan MCMC- eli Markovin ketju Monte Carlo -menetelmällä (Gelman ym., 2014) posteriorijakaumat, ja sitä varten simuloidaan neljä toisistaan riippumatonta ketjua, joiden aloitusarvot arvotaan satunnaisesti priorijakaumista. Jokaisen ketjun alussa on 1000 iteraation mittainen adaptaatiovaihe, minkä jälkeen on 4000 iteraation mittainen lämmitysjakso. Näiden vaiheiden jälkeen posteriorijakaumia simuloidaan kussakin ketjussa 4000 iteraation verran. Ketjujen konvergenssia tarkastellaan kuvien sekä Gelmanin ja Rubinin (1992) esittelemän $\hat{R}$-tunnusluvun (potential scale reduction) sekä sen monimuuttujaversion avulla (Brooks ja Gelman, 1998). Koko mallin sovittamiseen kuluu tavallisella PC-koneella noin 3.5 minuuttia. Käytetty JAGS-koodi on liitteessä 1.

Luvun 3.3.1 mukaisen mallin regressiokertoimien posteriorikeskiarvot, -keskihajonnat ja 95 % posteriorivälit on esitetty taulukossa 6. Kertoimien estimointi näyttää $\hat{R}$-tunnuslukujen ja kuvatarkastelujen (Liite 2) perusteella onnistuvan hyvin. Kertoimien posteriorijakaumat muistuttavat normaalijakaumia. Painoindeksiin ja sosioekononomiseen luokkaan *johtaja* liittyvät regressiokertoimet näyttävät eroavan nollasta. Painoindeksin kohoaminen yhdellä yksiköllä kasvattaa riskiryhmään kuulumisen todennäköisyyttä suhteessa terveenä olemisen todennäköisyyteen noin $e^{0.129} = 1.14$-kertaiseksi. Verrattuna sosioekonomisen luokan *muu* edustajaan johtajana työskentelevän todennäköisyys kuulua riskiryhmään suhteessa todennäköisyyteen olla terve vaikuttaa olevan $e^{-0.572} = 0.56$-kertainen. Myös toimistotyön tekeminen, ammattiurheilijatausta ja suurempi liikunnallinen aktiivisuus vapaa-ajalla näyttävät liittyvän pienempään todennäköisyyteen kuulua riskiryhmään. Ikä puolestaan vaikuttaa liittyvän korkeampaan todennäköisyyteen kuulua riskiryhmään.

Taulukko 6: *Riskiryhmään vuonna 2008 kuulumisen todennäköisyyteen, l, vaikuttavat regressiokertoimien estimaatit, posteriorikeskihajonnat ja 95 %:n posteriorivälit. Viimeisessä sarakkeessa on konvergenssista kertova tunnusluku* $\hat{R}$.

Kerroin	Keskiarvo	Keskihajonta	2.5 %	97.5 %	$\hat{R}$
β_{l_0}	0.509	0.213	0.093	0.925	1.001
$\beta_{l_{ika}}$	0.027	0.017	-0.006	0.061	1.001
$\beta_{l_{BMI}}$	0.129	0.038	0.055	0.205	1.000
$\beta_{l_{MET}}$	-0.001	0.004	-0.008	0.006	1.000
$\beta_{l_{urh}}$	-0.088	0.224	-0.529	0.355	1.001
$\beta_{l_{toim}}$	-0.357	0.241	-0.825	0.115	1.001
$\beta_{l_{joht}}$	-0.572	0.266	-1.096	-0.047	1.001

Mallin mukaisesti noin 65-vuotiaan sosioekonomista luokkaa *muu* edustavan miehen, jonka painoindeksi on noin 26 ja MET-arvo noin 31, todennäköisyys kuulua riskiryhmään on likimain 62 prosenttia. Se osa aineistosta, jolle malli on sovitettu, sisältää noin 53 prosenttia riskiryhmään kuuluvia.

Verrataan vielä urheilijan ja verrokin riskiryhmään kuulumisen todennäköisyyksiä, kun verrattavat henkilöt ovat muuten taustatekijöiltään samanlaisia. Valitaan aluksi vertailtavaksi 58-vuotiaat johtajat, joiden painoindeksi on 22 ja MET-arvo 18 eli kyseiset miehet ovat aineiston keskimääräistä tasoa nuorempia ja hieman kevyempiä, mutta vähemmän liikkuvia. Tällöin urheilijan todennäköisyys kuulua riskiryhmään on 30 prosenttia ja verrokin 32 prosenttia. Kun verrattavaksi valitaan keskimääräistä enemmän painavat ja vähemmän liikkuvat 72-vuotiaat sosioekonomisen luokan *muu* edustajat, joiden painoindeksi on 30 ja MET-arvo 15, on mallin mukaan urheilijan todennäköisyys kuulua riskiryhmään noin 76 prosenttia ja verrokin noin 77 prosenttia.

4.2 Malliin 2 liittyvät posteriorijakaumat

Käydään seuraavaksi läpi, miten taustamuuttujien vaikutukset ja siirtymätodennäköisyydet estimoituvat mallin 2 tapauksessa. Kovariaateista ikä, painoindeksi, ja MET on keskistetty ennen analyysien tekemistä tutkittavien vuosien yhteiselle

keskimääräiselle tasolle. Riskiryhmään kuuluvat henkilöt sisällytetään terveisiin, jolloin mahdollisia tiloja on kolme. Kovariaattien puuttuva tieto täydennetään nyt kuten edellä eli

$$BMI \sim N(0, 10),$$

$$MET \sim N(0, 700),$$

$$toim \sim \text{Bernoulli}(0.35),$$

ja

$$joht \sim \text{Bernoulli}(0.25),$$

kun jatkuvien muuttujien kohdalla on tehty keskistys ja normaalijakauman hajontaparametrina on varianssi. Kaikille regressiokertoimille asetetaan prioriksi normaalijakauma $N(0, 2)$, missä hajontaparametrina on varianssi.

Myös tämä malli sovitetaan R- ja JAGS-ohjelmien avulla. Regressiokertoimien posteriorijakaumat estimoidaan MCMC-menetelmällä simuloimalla neljä toisistaan riippumatonta ketjua, joiden aloitusarvot arvotaan satunnaisesti priorijakaumista. Jokaisen ketjun aluksi simuloidaan 1000 iteraation mittainen adaptaatiovaihe, minkä jälkeen simulointia jatketaan 2000 iteraation mittaisella lämmitysjaksolla. Näiden vaiheiden jälkeen simuloidaan kussakin ketjussa 2000 iteraation verran posteriorijakaumia. Ketjujen konvergenssia tarkastellaan kuvien sekä $\hat{R}$-tunnusluvun ja sen monimuuttujaversion avulla. Koko mallin sovittamiseen kuluu tavallisella PC-koneella noin 20 tuntia. Käytetty JAGS-koodi on liitteessä 3.

Sovitetun mallin estimoidut regressiokertoimet keskihajontoineen ja 95 prosentin posteriorivälеineen on kuvattu taulukossa 7. Kuvatarkasteluiden (Liite 4) ja $\hat{R}$-tunnuslukujen perusteella kaikkiin siirtymiin liittyvät kertoimet vaikuttavat estimoituvan hyvin. Posteriorijakaumat muistuttavat normaalijakaumia. Kun kertoimeen liittyvän muuttujan arvo kasvaa yksikön, kertoimen eksponoitu arvo kertoo, kuinka paljon todennäköisyys tilasiirtymälle, johon kerroin liittyy, muuttuu suhteessa samassa tilassa pysymisen todennäköisyyteen. Tulkittaessa yksittäistä

kerrointa oletetaan kaikkien muiden taustekijöiden pysyvän vakioina.

Tulkitaan ensin terveen henkilön kuolemiseen liittyvien regressiokertoimien posteriorikeskiarvoja. Referenssihenkilönä on noin 66.7-vuotias terve mies, jonka painoindeksi on 26.4 ja MET-arvo 25.2. Hän ei ole taustaltaan huippu-urheilija ja kuuluu sosioekonomiseen luokkaan *muu*. Todennäköisyys, että tällainen terve referenssihenkilö kuolee, suhteessa terveenä pysymiseen, on $e^{-3.402} = 0.03$-kertainen. Muiden taustatekijöiden pysyessä samoina ikääntyminen vuoden verran muuttaa kuolemisen riskin suhteessa terveenä pysymiseen $e^{0.098} = 1.1$-kertaiseksi ja painoindeksin kasvaminen yhdellä yksiköllä $e^{-0.026} = 0.97$-kertaiseksi. MET-arvon kasvaminen yhdellä yksiköllä puolestaan pienentää kuolemisen riskiä suhteessa terveenä pysymiseen $e^{-0.010} = 0.99$-kertaiseksi. Huippu-urheilijatausta näyttää tässä pienentävän kuolemisen riskin suhteessa terveenä pysymiseen noin $e^{-0.253} = 0.78$-kertaiseksi. Terveen toimistotyöntekijän todennäköisyys kuolla on $e^{-0.157} = 0.85$-kertainen suhteessa terveenä pysymiseen. Terveen johtajan todennäköisyys kuolla on puolestaan $e^{-0.221} = 0.80$-kertainen suhteessa terveenä pysymiseen. Toimistotyöntekijöiden ja johtajien riski kuolla vaikuttaa olevan siis hieman pienempi kuin sosioekonomiselta taustaltaan muiden. Painoindeksin, toimistotyöntekijöiden ja johtajien regressiokertoimien 95 prosentin posteriorivälit tosin sisältävät nollan.

Käydään sitten läpi terveen henkilön tyypin 2 diabetekseen sairastumiseen liittyvien regressiokertoimien posteriorikeskiarvoja. Edellä kuvatun terveen referenssihenkilön todennäköisyys tyypin 2 diabetekseen sairastumiseen on $e^{-4.434} = 0.01$-kertainen suhteessa terveenä pysymiseen. Vuoden verran vanheneminen kasvattaa riskiä sairastua $e^{0.062} = 1.06$-kertaiseksi verrattuna todennäköisyyteen pysyä terveenä. Painoindeksin kasvulla on sairastumisriskiä kasvattava vaikutus. Kun painoindeksi kasvaa yhdellä yksiköllä, kasvaa riski sairastumiselle $e^{0.170} = 1.19$-kertaiseksi suhteessa terveenä pysymisen todennäköisyyteen. MET-arvon nousemisella näyttää olevan sairastumisriskiä hieman kasvattava vaikutus, mutta käytännössä sairastumisen riski on $e^{0.002} = 1.00$-kertainen verrattuna todennäköisyyteen pysyä terveenä. Huippu-urheilijataustaisen miehen sairastumisen todennäköisyys on $e^{-0.196} = 0.82$-kertainen suhteessa terveenä pysymisen todennäköisyyteen

eli huippu-urheilijaura pienentää myös sairastumisen todennäköisyyttä. Sosioekonomisen luokan ollessa jokin toinen kuin *muu* todennäköisyys sairastua pienenee. Toimistotyöntekijän todennäköisyys sairastumiselle on $e^{-0.422} = 0.66$-kertainen ja johtajan $e^{-0.489} = 0.61$-kertainen suhteessa terveenä pysymisen todennäköisyyteen. MET-arvoon ja urheilijastatukseen liittyvien regressiokertoimien 95 prosentin posteriorivälit sisältävät nollan.

Tutkitaan vielä taustamuuttujien vaikutusta tyypin 2 diabeetikon kuolemisen todennäköisyyteen. Nyt referenssihenkilö on muuten kuten edellä, mutta hän ei ole terve vaan hänellä on tyypin 2 diabeteksen diagnoosi. Tällaisen referenssihenkilön kuolemisen todennäköisyys on $e^{-3.056} = 0.05$-kertainen suhteessa elävänä diabeetikkona pysymiseen. Ikääntyminen vuoden verran kasvattaa kuolemisen riskin $e^{0.079} = 1.08$-kertaiseksi suhteessa diabeetikon todennäköisyyteen pysyä elossa. Painoindeksin nousu yhdellä yksiköllä puolestaan kasvattaa kuolemisen todennäköisyyden $e^{0.048} = 1.05$-kertaiseksi suhteessa elossa säilymisen todennäköisyyteen. Nyt MET-arvon kasvu näyttää hieman pienentävän kuolemisen riskiä, mutta käytännössä yhden yksikön kasvu MET-arvossa säilyttää kuolemisen todennäköisyyden $e^{-0.003} = 1.00$-kertaisena verrattuna todennäköisyyteen pysyä elossa. Nyt huippu-urheilijataustalla vaikuttaa olevan kuolemisen todennäköisyyttä kasvattava vaikutus. Henkilön ollessa huippu-urheilija todennäköisyys kuolla on $e^{0.072} = 1.07$-kertainen suhteessa elossa säilymisen todennäköisyyteen, kun verrattavien miesten taustat eroavat ainoastaan urheilijastatukseltaan. Diabetesta sairastavan toimistotyöntekijän todennäköisyys kuolla on $e^{-0.349} = 0.71$-kertainen ja johtajan $e^{-0.623} = 0.54$-kertainen suhteessa todennäköisyyteen pysyä elossa, kun verrattavat henkilöt eroavat vain sosioekonomiselta luokaltaan. Näistä regressiokertoimista ikään liittyvä on kuitenkin ainoa, joka ei sisällä nollaa 95 % posteriorivälillään.

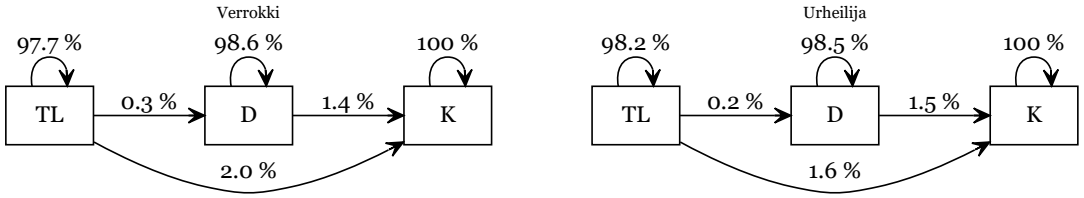
Taulukko 7: *Mallin 2 regressiokertoimien estimaatit, posteriorikeskihajonnat ja 95 %:n posteriorivälit. Viimeisessä sarakkeessa on konvergenssista kertova tunnusluku* $\hat{R}$.

Siirtymä	Kerroin	Keskiarvo	Keskihajonta	2.5 %	97.5 %	$\hat{R}$
TL → K	β_{p_0}	-3.402	0.081	-3.565	-3.245	1.005
	$\beta_{p_{ika}}$	0.098	0.004	0.090	0.107	1.001
	$\beta_{p_{BMI}}$	-0.026	0.015	-0.056	0.004	1.001
	$\beta_{p_{MET}}$	-0.010	0.002	-0.015	-0.006	1.001
	$\beta_{p_{urh}}$	-0.253	0.093	-0.434	-0.068	1.004
	$\beta_{p_{toim}}$	-0.157	0.099	-0.349	0.035	1.001
	$\beta_{p_{joht}}$	-0.221	0.116	-0.447	0.003	1.000
TL → D	β_{q_0}	-4.434	0.143	-4.721	-4.164	1.001
	$\beta_{q_{ika}}$	0.062	0.009	0.044	0.080	1.000
	$\beta_{q_{BMI}}$	0.170	0.021	0.127	0.211	1.001
	$\beta_{q_{MET}}$	0.002	0.003	-0.005	0.008	1.001
	$\beta_{q_{urh}}$	-0.196	0.172	-0.531	0.147	1.001
	$\beta_{q_{toim}}$	-0.422	0.191	-0.808	-0.057	1.002
	$\beta_{q_{joht}}$	-0.489	0.228	-0.945	-0.053	1.001
D → K	β_{r_0}	-3.056	0.203	-3.463	-2.671	1.001
	$\beta_{r_{ika}}$	0.079	0.014	0.051	0.107	1.001
	$\beta_{r_{BMI}}$	0.048	0.025	-0.003	0.096	1.000
	$\beta_{r_{MET}}$	-0.003	0.007	-0.016	0.010	1.002
	$\beta_{r_{urh}}$	0.072	0.225	-0.369	0.515	1.001
	$\beta_{r_{toim}}$	-0.349	0.234	-0.807	0.102	1.001
	$\beta_{r_{joht}}$	-0.623	0.346	-1.343	0.024	1.001

Kaavan 2 mukaisesti voidaan edelleen laskea varsinaisia siirtymätodennäköisyyksiä. Tutkitaan ensin todennäköisyyksiä, jotka koskevat edellä kuvattua referenssihenkilöä, joka on aineiston mukaisesti keskimääräisen ikäinen eli 66.7-vuotias ja painoinen, mikä vastaa painoindeksin arvoa 26.4. Myös hänen vapaa-ajan liikunnan raskaustasonsa on keskimääräistä tasoa MET-arvon ollessa 25.2. Hän ei ole ollut aikaisemmin huippu-urheilija ja kuuluu sosioekonomiseen luokkaan *muu*. Sovitettu malli estimoi taustatekijöiltään juuri kuvatun terveen henkilön todennäköisyydeksi sairastua tyypin 2 diabetekseen yhden vuoden aikana (q) noin 1 prosenttia. Saman terveen henkilön todennäköisyys kuolla (p) on puolestaan noin 3 prosenttia. Todennäköisyys pysyä terveenä ja elossa on siten noin 96 prosent-

tia. Mikäli taustatekijät pysyvät samoina, mutta henkilön lähtötila on tyypin 2 diabeetikko, on kuolemisen todennäköisyys (r) noin 4 prosenttia ja elossa pysymisen todennäköisyys 96 prosenttia. Taustatekijöiden muuttuminen kasvattaa ja pienentää siirtymätodennäköisyyksiä vetokerrointen kautta kuvatulla tavalla. Esimerkiksi ikääntyminen kasvattaa sairastumisen ja kuolemisen todennäköisyyttä ja erityisesti johtajien riski sairastua ja kuolla vaikuttaa olevan pienempi kuin muiden sosioekonomisten luokkien edustajien.

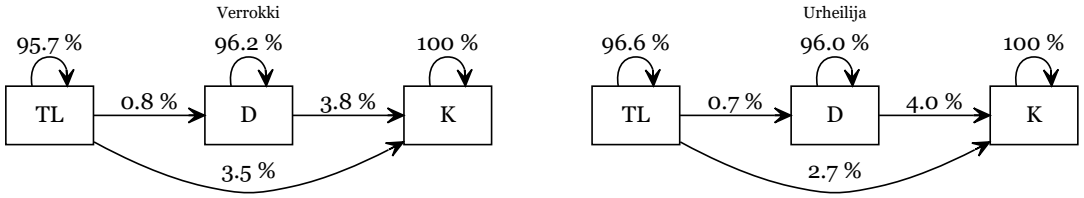
Havainnollistetaan siirtymätodennäköisyyksiä erilaisilla taustamuuttujien arvoilla vielä kolmen esimerkkitapauksen avulla. Ensimmäisessä tapauksessa vartailtavana on kaksi miestä, joiden ainoana erona on, että toinen on entinen huippuurheilija ja toinen verrokki. Miehet ovat 58-vuotiaita ja työskentelevät johtajina. Kummankin painoindeksi on 22 ja MET-arvo 18 eli molemmat arvot ovat analysoitavan aineiston keskimääräistä painoindeksiä ja MET-arvoa pienempiä. Kuvassa 3 ovat mallin 2 mukaiset siirtymätodennäköisyydet eri tilojen välillä tällaisille henkilöille. Verrokin ja urheilijan siirtymätodennäköisyydet ovat hyvin samankaltaisia. Suurin ero on, että terveen verrokin todennäköisyys kuolla on 0.4 prosenttiyksikköä suurempi kuin terveen urheilijan.



Kuva 3: *Tilanteen 1 esimerkkihenkilöiden taustatekijöitä vastaavat siirtymätodennäköisyydet verrokille vasemmalla ja urheilijalle oikealla.*

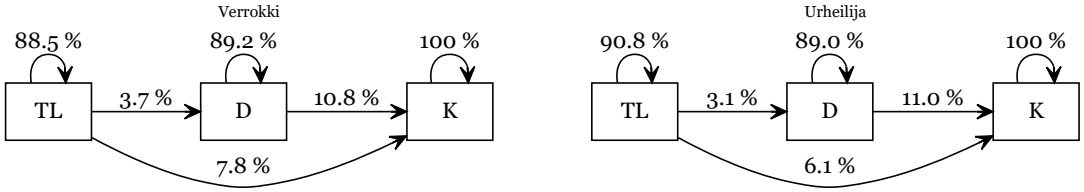
Toisessa esimerkkitapauksessa verrattavat miehet ovat 65-vuotiaita toimistotyöntekijöitä. Molempien painoindeksi on 26 ja MET-arvo 25, jotka vastaavat likimain aineiston keskimääräistä tasoa. Kuvassa 4 on esitetty tällaisen verrokin ja urheilijan mallin 2 mukaiset siirtymätodennäköisyydet. Edelleen todennäköisyyksissä urheilijan ja verrokin välillä on melko pieniä eroja, suurimpana terveen verrokin 0.8 prosenttiyksikköä suurempi kuolintodennäköisyys kuin urheilijan. Ver-

rattuna ensimmäiseen esimerkkitapaukseen todennäköisyys pysyä lähtötilassa on nyt pienempi kaikkien tilojen kohdalla.



Kuva 4: *Tilanteen 2 esimerkkihenkilöiden taustatekijöitä vastaavat siirtymätodennäköisyydet verrokille vasemmalla ja urheilijalle oikealla.*

Kolmannessa esimerkkitapauksessa verrataan kahta 72-vuotiasta miestä, jotka edustavat sosioekonomista luokkaa *muu*. Molempien painoindeksi on tässä tapauksessa 30 ja MET-arvo 15. Molempien painoindeksi on siis keskimääräistä suurempi ja MET-arvo keskimääräistä pienempi analysoitavan aineiston tasolla. Kuvaan 5 on merkitty mallin 2 mukaiset siirtymätodennäköisyydet näille esimerkin mukaisille miehille. Nyt terveen verrokin todennäköisyys kuolla on 1.7 prosenttiyksikköä suurempi kuin urheilijan. Tässä tapauksessa todennäköisyydet pysyä lähtötilassa ovat suurin piirtein 90 prosentin tasoa eli useamman prosenttiyksikön pienempiä kuin edellisissä tapauksissa.



Kuva 5: *Tilanteen 3 esimerkkihenkilöiden taustatekijöitä vastaavat siirtymätodennäköisyydet verrokille vasemmalla ja urheilijalle oikealla.*

5 Pohdinta

Tämän tutkielman tarkoituksena on arvioida siirtymätodennäköisyyksiä liittyen tyypin 2 diabeteksen etenemiseen ja taustatekijöiden vaikutusta näihin todennäköisyyksiin. Tutkittavaksi valittiin viisi taustamuuttujaa, joista MET ja urheilustatus liittyvät liikuntaan ja painoindeksin voi ajatella kuvaavan laajemmin yksilön elämäntapoja. Ikä ja sosioekonominen status kertovat muuten henkilön taustasta.

Aineistoon sovitettiin tilasiirtymämalli, joka ei ota erikseen huomioon piilevää tyypin 2 diabetesta tai sen esiastetta, joita kutsutaan tässä työssä yhteisnimityksellä riskiryhmä. Kyseisen mallin mukaan sosioekonomiselta statukseltaan luokkaan *muu* kuuluvan miehen, jonka ikä, painoindeksi ja vapaa-ajanliikunnan raskaustaso ovat aineiston tasolla keskimääräisiä ja joka ei ole ollut huippu-urheilija, todennäköisyys sairastua tyypin 2 diabetekseen yhden vuoden aikana on noin yksi prosentti. Mallin mukaan todennäköisyys, että vastaava terve henkilö kuolee, on noin kolme prosenttia. Diabeetikon kuolemisen riski on vielä hieman suurempi, noin neljä prosenttia.

Mallin mukaan ikääntyminen on yhteydessä sekä diabetekseen sairastumisen että kuolemisen suurempiin riskeihin, mikä vaikuttaa luonnolliselta. Huippu-urheilijatausta näyttää mallin ja aineiston perusteella liittyvän terveen henkilön kuolemisen ja tyypin 2 diabetekseen sairastumisen pienempään riskiin ja suurempaan tyypin 2 diabeetikon kuolemistodennäköisyyteen. Suuret painoindeksin arvot näyttävät olevan yhteydessä terveen henkilön kohonneeseen sairastumisen todennäköisyyteen, mutta yhteyttä muihin siirtymätodennäköisyyksiin niillä ei näytä olevan. Johtajana työskenteleminen vaikuttaa olevan yhteydessä pienempään terveiden riskiin sairastua tyypin 2 diabetekseen kuin toisten sosioekonomisten luokkien edustajien. Tämä malli näyttää toimivan melko hyvin, vaikka hieman yllättäen liikunnan vaikutukset näyttävät olevan suurelta osin analysoitavassa aineistossa sattumaa. Mallia tulkitessa on syytä kiinnittää huomiota siihen, että kyseessä on ennustemalli eikä kausaalimalli, jolloin ei voida puhua syistä tai seurauksista mallin tuloksien

suhteen.

Riskiryhmään kuulumisen todennäköisyyttä vuonna 2008 ehdolla vuoden 2001 taustamuuttujien arvot estimoivalla mallilla saatiin riskiryhmään kuulumisen todennäköisyydeksi aineiston keskimääräistä tasoa edustavalle henkilölle noin 62 prosenttia. Kohonnut painoindeksi näyttää liittyvän suurempaan todennäköisyyteen kuulua riskiryhmään ja johtajana työskenteleminen pienempään todennäköisyyteen kuulua riskiryhmään. Myös ikääntyminen näyttää kasvattavan riskiryhmään kuulumisen todennäköisyyttä, mikä vaikuttaa luonnolliselta. Iän vaikutus saattaa tosin olla analysoitavan aineiston mukaan sattumaa.

Tilasiirtymämallissa ei sallittu palaavia siirtymiä edellisiin tiloihin. Tämä oletus tehtiin laskennallisen yksinkertaisuuden vuoksi, mutta sen realistisuutta voi kritisoida etenkin prediabeteksen kohdalla. Omilla elämäntavoillaan on kuitenkin mahdollista vaikuttaa tyypin 2 diabetekseen sairastumiseen (Scobie ja Samaras, 2009, 24-29). Sovitetun mallin vahvuutena voisi pitää vuosittaisten todennäköisyyksien estimoimista ottamalla huomioon tutkimuskertojen välillä vallitsevat eri mittaiset ajanjaksot.

Käytettävän aineiston saattaminen analysoitavaan muotoon vaati runsaasti aikaa. Henkilöt, joilta puuttui muuttujien arvoja vain joidenkin muuttujien kohdalla, haluttiin sisällyttää tutkimukseen, mikä vaati puuttuvan tiedon imputoimista. Puuttuvat ikätiedot saatiin yksinkertaisin yhteen- ja vähennyslaskuin, kun ikä tiedettiin edes yhtenä tutkimusvuotena. Analyysiin ei otettu mukaan yhtään henkilöä, jonka ikää ei olisi tiedetty lainkaan. Muiden taustatietojen täydentäminen tehtiin arpomalla satunnaisesti kunkin muuttujan noudattamista, aineistosta arvioiduista, jakaumista, mikä ei ole kaikkien muuttujien kohdalla paras mahdollinen tapa. Painoindeksin voisi olettaa olevan yhteydessä yhdellä tutkimuskerralla edellisen ja seuraavan tutkimuskerran havaintoihin. Sama voisi olla järkevä oletus myös liikunnan rasittavuudesta kertovan MET-muuttujan kohdalla. Näistä muuttujista puuttuvat tiedot voisi siten olla realistisempaa arvioida perustuen nimenomaan kyseisen henkilön muihin saman muuttujan havaintoarvoihin, mikäli henkilö on vastannut aikaisemmin tai myöhemmin.

Aineisto sisältää myös monia muita mahdollisia taustamuuttujia, joita ei tässä työssä otettu huomioon. Esimerkiksi tupakoinnin (Xu ym., 2012; Okwechime ym., 2016) ja alkoholinkululutuksen (Xu ym., 2012) yhteyttä tyypin 2 diabetekseen sairastumiseen on tutkittu aiemmin. Nämä muuttujat olisivat saatavilla tästäkin aineistosta. Erityisesti liikunnallisuutta silmällä pitäen kerätty aineisto pitää sisällään myös tiedon entisten huippu-urheilijoiden edustamista lajeista. Lajitiedon avulla voitaisiin mahdollisesti selvittää eri tyyppisen, kuten voima- tai kestävyyslajin urheilun vaikutusta siirtymätodennäköisyyksiin. Nyt lajin, tupakoinnin tai alkoholin merkitystä ei kuitenkaan tutkittu, sillä alustavissa tarkasteluissa saatujen tulosten perusteella mukaan valittiin vain tärkeimmiltä vaikuttavat muuttujat.

Koska aineisto lähtökohtaisesti koostuu mieshuippu-urheilijoista ja heille valituista verrokkihenkilöistä, ei saatuja tuloksia voida yleistää koskemaan kaikkia suomalaisia, eikä edes kaikkia suomalaisia miehiä. Tutkittavien taustat ovat liian spesifisti rajattuja tällaisten yleistysten tekemiseen. Tämä on otettava huomioon tulkintoja mietittäessä.

Viitteet

Alberti, K.G.M.M., Zimmet, P. ja Shaw, J. Metabolic syndrome - a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabetic Medicine*, 23(5):469–480, 2006.

Brooks, S.P. ja Gelman, A. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

Dobson, A.J. ja Barnett, A.G. *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 3. painos, 2008.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. ja Rubin, D.B. *Bayesian Aata Analysis*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, Boca Raton, 3 painos, 2014.

Gelman, A. ja Rubin, D.B. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472, 1992.

Honeycutt, A.A., Boyle, J.P., Broglio, K.R., Thompson, T.J., Hoerger, T.J., Geiss, L.S. ja Venkat Narayan, K. A Dynamic Markov Model for Forecasting Diabetes Prevalence in the United States through 2050. *Health Care Management Science*, 6(3):155–164, 2003.

Hosmer, D.W., Lemeshow, S. ja Sturdivant, R.X. *Applied Logistic Regression*. Wiley Series in Probability and Statistics. Wiley, 3. painos, 2013.

Kujala, U.M., Kaprio, J., Sarna, S. ja Koskenvuo, M. Relationship of Leisure-Time Physical Activity and Mortality: The Finnish Twin Cohort. *JAMA*, 279(6):440–444, 1998.

Kujala, U.M., Marti, P., Kaprio, J., Hernelahti, M., Tikkanen, H. ja Sarna, S.

Occurrence of Chronic Disease in Former Top-Level Athletes. *Sports Medicine*, 33(8):553–561, 2003.

Kujala, U.M., Peltonen, M., Laine, M.K., Kaprio, J., Heinonen, O.J., Sundvall, J., Eriksson, J.G., Jula, A., Sarna, S. ja Kainulainen, H. Branched-Chain Amino Acid Levels Are Related with Surrogates of Disturbed Lipid Metabolism among Older Men. *Frontiers in Medicine*, 3:57, 2016.

Okwechime, I.O., Roberson, S. ja Odoi, A. Prevalence and Predictors of Pre-Diabetes and Diabetes among Adults 18 Years or Older in Florida: A Multinomial Logistic Modeling Approach. *PLOS ONE*, 10(12), 2016.

Plummer, M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Julkaisussa *3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, 124(125.10):1–10, Vienna, Austria, 2003.

Plummer, M. *rjags: Bayesian Graphical Models using MCMC*, 2019. R package version 4-9. https://CRAN.R-project.org/package=rjags.

Plummer, M., Best, N., Cowles, K. ja Vines, K. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11, 2006. https://journal.r-project.org/archive/.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. http://www.R-project.org/.

Roberts, C.K., Hevener, A.L. ja Barnard, R.J. Metabolic Syndrome and Insulin Resistance: Underlying Causes and Modification by Exercise Training. *Comprehensive Physiology*, 3:1–58, 2013.

Sarna, S., Sahi, T., Koskenvuo, M. ja Kaprio, J. INCREASED LIFE EXPECTANCY OF WORLD CLASS MALE-ATHLETES. *Medicine and Science in Sports and Exercise*, 25:237–244, 1993.

Saukkonen, T. *Prediabetes and associated cardiovascular risk factors: A prospective cohort study among middle-aged and elderly Finns.* Väitöskirja, Oulu, 2012.

Scobie, I. ja Samaras, K. *Fast Facts: Diabetes Mellitus.* Health Press Limited, Abingdon, 3. painos, 2009.

Suomalaisen Lääkäriseuran Duodecimin, Suomen Sisätautilääkärien yhdistyksen ja Diabetesliiton Lääkärineuvoston asettama työryhmä. Tyypin 2 diabetes. Käypä hoito -suositus. Helsinki: Suomalainen Lääkäriseura Duodecim, 2018. Luettu 31.3.2019. Saatavilla internetissä: www.kaypahoito.fi.

Suomen Diabetesliitto ry. Tilastotietoa diabeteksesta. 2019. Luettu 5.4.2019. https://www.diabetes.fi/diabetes/yleista_diabeteksesta/tilastotietoa.

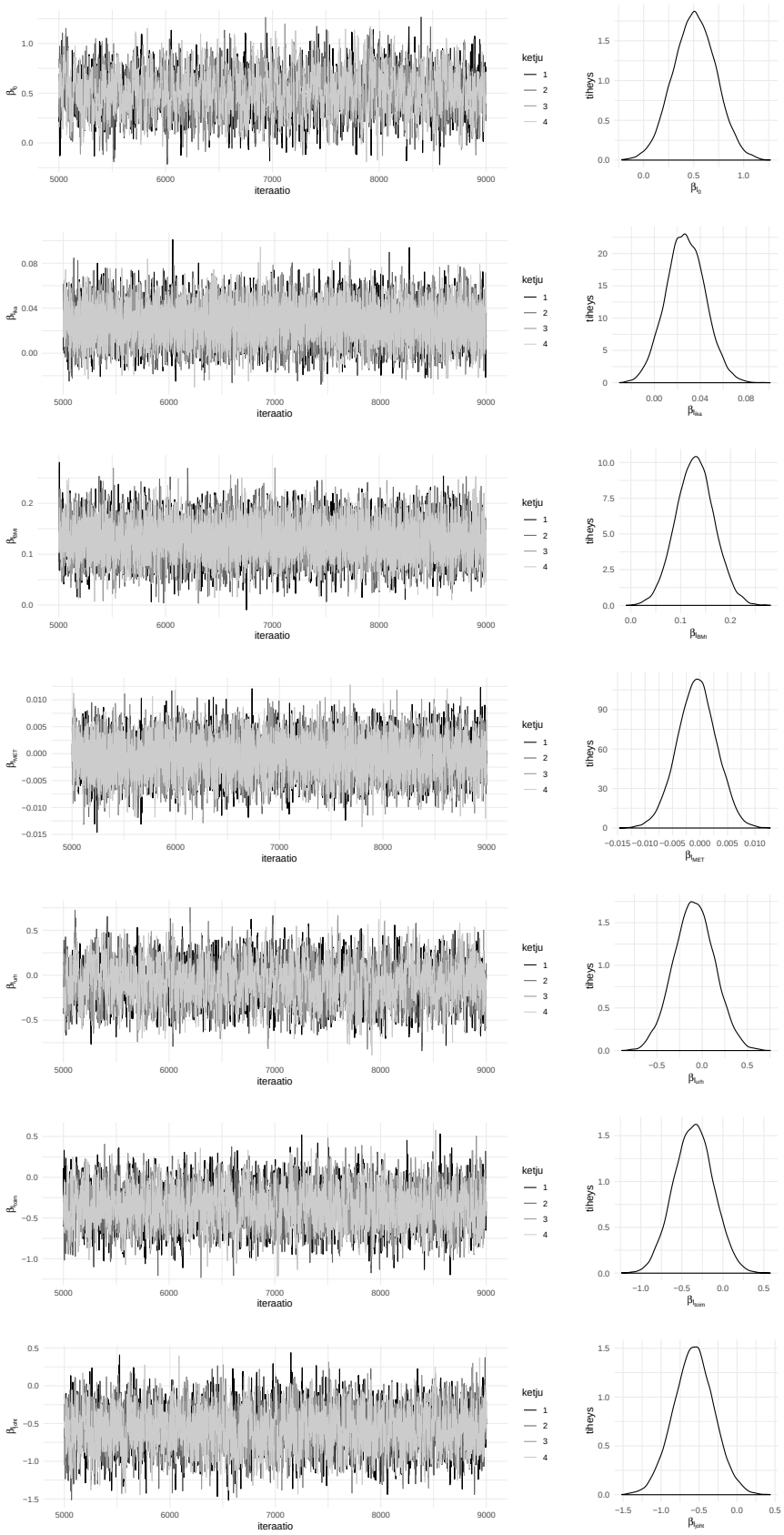
Xu, W., Xu, Z., Jia, J., Xie, Y., Wang, H.X. ja Qi, X. Detection of Prediabetes and Undiagnosed Type 2 Diabetes: A Large Population-Based Study. *Canadian Journal of Diabetes*, 36(3):108–113, 2012.

Liitteet

Liite 1. JAGS-koodi riskiryhmään kuulumisen todennäköisyyden mallille

```
## Sovitetaan malli riskiryhmaan vuonna 2008 kuulumisen
## todennakoisyydelle.
logistinen_malli <- "model{
  # Vasteet Bernoullijakaumasta parametrilla l
  for(i in 1:N){
    y[i] ~ dbern(l[i])
    logit(l[i]) <- inprod(beta_l, X[i, ])
  }

  # Priorit regressiokertoimille
  for(j in 1:p){
    beta_l[j] ~ dnorm(0, 1 / 2)
  }

  # Priorit puuttuville kovariaateille
  for(n in 1:N){
    X[n, 3] ~ dnorm(0, 1 / 10)  # BMI
    X[n, 4] ~ dnorm(0, 1 / 700) # MET
    X[n, 6] ~ dbern(0.35)       # toimistotyontekija
    X[n, 7] ~ dbern(0.25)       # johtaja
  }
}"

# Mallin sovitus
# Datana vasteet y, kovariaattien matriisi X,
# matriisin X sarakkeiden lukumaara p, havaintojen
# lukumaara N.

model <- jags.model(textConnection(logistinen_malli), data = par,
  n.chains = 4)
update(model, n.iter = 4000)
samp <- coda.samples(model, variable.names = c("beta"),
  n.iter = 4000)
```

Liite 2. Riskiryhmään kuulumisen todennäköisyyden mallin sovitukseen liittyvät kuvat



Kuva 1: *Riskiryhmään kuulumisen todennäköisyyteen l liittyvien regressiokertoimien ketjut vasemmalla ja tiheyskuvaajat ketjuista lasketuille kertoimien estimaateille oikealla. Kyseinen l vastaa vuoden 2008 riskiryhmään kuulumisen todennäköisyyttä*

Liite 3. JAGS-koodi mallille 2

```
## Sovitetaan malli 2 ilman erillista riskiryhmaa.
malli <- "model{
  ## Siirtyma terveista 85 -> 95
  for(i in 1:N_terve85) {
    # Vasteet kategorisesta jakaumasta parametrilla par_terve85
    y[i] ~ dcat(par_terve85[i, ])

    # Siirtyma terveesta terveeksi
    par_terve85[i, 1] <- (1 - p[i] - q[i])^10

    # Siirtyma terveesta diagnosoiduksi
    par_terve85[i, 2] <- q[i] * (1 - r[i])^9 +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^8 +
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^7 +
      (1 - p[i] - q[i])^3 * q[i] * (1 - r[i])^6 +
      (1 - p[i] - q[i])^4 * q[i] * (1 - r[i])^5 +
      (1 - p[i] - q[i])^5 * q[i] * (1 - r[i])^4 +
      (1 - p[i] - q[i])^6 * q[i] * (1 - r[i])^3 +
      (1 - p[i] - q[i])^7 * q[i] * (1 - r[i])^2 +
      (1 - p[i] - q[i])^8 * q[i] * (1 - r[i]) +
      (1 - p[i] - q[i])^9 * q[i]

    # Siirtyma terveesta kuolleeksi
    par_terve85[i, 3] <- p[i] + (1 - p[i] - q[i]) * p[i] +
        (1 - p[i] - q[i])^2 * p[i] + (1 - p[i] - q[i])^3 * p[i] +
        (1 - p[i] - q[i])^4 * p[i] + (1 - p[i] - q[i])^5 * p[i] +
        (1 - p[i] - q[i])^6 * p[i] + (1 - p[i] - q[i])^7 * p[i] +
        (1 - p[i] - q[i])^8 * p[i] + (1 - p[i] - q[i])^9 * p[i] +
      q[i] * r[i] + (1 - p[i] - q[i]) * q[i] * r[i] +
        (1 - p[i] - q[i])^2 * q[i] * r[i] +
        (1 - p[i] - q[i])^3 * q[i] * r[i] +
        (1 - p[i] - q[i])^4 * q[i] * r[i] +
        (1 - p[i] - q[i])^5 * q[i] * r[i] +
        (1 - p[i] - q[i])^6 * q[i] * r[i] +
        (1 - p[i] - q[i])^7 * q[i] * r[i] +
        (1 - p[i] - q[i])^8 * q[i] * r[i] +
      q[i] * (1 - r[i]) * r[i] +
        (1 - p[i] - q[i]) * q[i] * (1 - r[i]) * r[i] +
        (1 - p[i] - q[i])^2 * q[i] * (1 - r[i]) * r[i] +
        (1 - p[i] - q[i])^3 * q[i] * (1 - r[i]) * r[i] +
        (1 - p[i] - q[i])^4 * q[i] * (1 - r[i]) * r[i] +
        (1 - p[i] - q[i])^5 * q[i] * (1 - r[i]) * r[i] +
```

```
      (1 - p[i] - q[i])^6 * q[i] * (1 - r[i]) * r[i] +
      (1 - p[i] - q[i])^7 * q[i] * (1 - r[i]) * r[i] +
    q[i] * (1 - r[i])^2 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^2 * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^2 * r[i] +
      (1 - p[i] - q[i])^3 * q[i] * (1 - r[i])^2 * r[i] +
      (1 - p[i] - q[i])^4 * q[i] * (1 - r[i])^2 * r[i] +
      (1 - p[i] - q[i])^5 * q[i] * (1 - r[i])^2 * r[i] +
      (1 - p[i] - q[i])^6 * q[i] * (1 - r[i])^2 * r[i] +
    q[i] * (1 - r[i])^3 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^3 * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^3 * r[i] +
      (1 - p[i] - q[i])^3 * q[i] * (1 - r[i])^3 * r[i] +
      (1 - p[i] - q[i])^4 * q[i] * (1 - r[i])^3 * r[i] +
      (1 - p[i] - q[i])^5 * q[i] * (1 - r[i])^3 * r[i] +
    q[i] * (1 - r[i])^4 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^4 * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^4 * r[i] +
      (1 - p[i] - q[i])^3 * q[i] * (1 - r[i])^4 * r[i] +
      (1 - p[i] - q[i])^4 * q[i] * (1 - r[i])^4 * r[i] +
    q[i] * (1 - r[i])^5 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^5 * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^5 * r[i] +
      (1 - p[i] - q[i])^3 * q[i] * (1 - r[i])^5 * r[i] +
    q[i] * (1 - r[i])^6 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^6 * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^6 * r[i] +
    q[i] * (1 - r[i])^7 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^7 * r[i] +
    q[i] * (1 - r[i])^8 * r[i]

    # Vuosittaiset todennakoisyydet p, q ja r
    p[i] <- p_[i] / (1 + p_[i] + q_[i])
    q[i] <- q_[i] / (1 + p_[i] + q_[i])
    r[i] <- r_[i] / (1 + r_[i])

    # Regressiomallit
    log(q_[i]) <- inprod(beta_q, X[i, ])
    log(r_[i]) <- inprod(beta_r, X[i, ])
    log(p_[i]) <- inprod(beta_p, X[i, ])
}

## siirtyma diagnosoiduista 85 -> 95
for(i in (N_terve85 + 1):(N_terve85 + N_diag85)) {
```

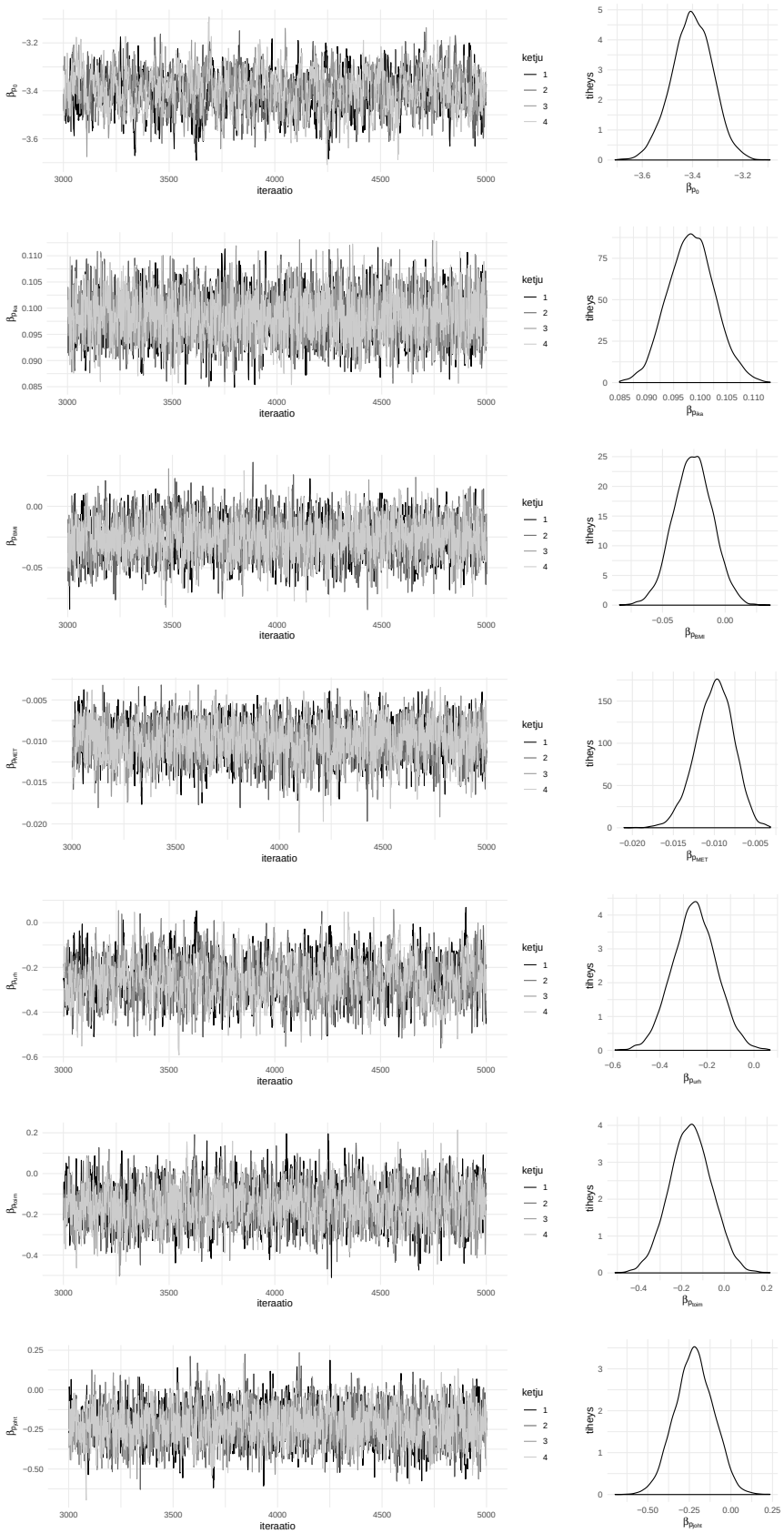
```
  # Vasteet Bernoulli-jakaumasta parametrilla r85
  y[i] ~ dbern(r85[i])

  # Siirtyma diagnosoidusta kuolleeksi
  r85[i] <- r[i] + (1 - r[i]) * r[i] +
    (1 - r[i])^2 * r[i] + (1 - r[i])^3 * r[i] +
    (1 - r[i])^4 * r[i] + (1 - r[i])^5 * r[i] +
    (1 - r[i])^6 * r[i] + (1 - r[i])^7 * r[i] +
    (1 - r[i])^8 * r[i] + (1 - r[i])^9 * r[i]

  # Vuosittainen todennakoisyys r ja sen malli
  logit(r[i]) <- inprod(beta_r, X[i, ])
}

## siirtyma terveista 95 -> 01
for(i in (N85 + 1):(N85 + N_terve95)) {
  # Vasteet kategorisesta jakaumasta parametrilla par_terve95
  y[i] ~ dcat(par_terve95[i, ])

  # Siirtyma terveesta terveeksi
  par_terve95[i, 1] <- (1 - p[i] - q[i])^6

  # Siirtyma terveesta diagnosoiduksi
  par_terve95[i, 2] <- q[i] * (1 - r[i])^5 +
    (1 - p[i] - q[i]) * q[i] * (1 - r[i])^4 +
    (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^3 +
    (1 - p[i] - q[i])^3 * q[i] * (1 - r[i])^2 +
    (1 - p[i] - q[i])^4 * q[i] * (1 - r[i]) +
    (1 - p[i] - q[i])^5 * q[i]

  # Siirtyma terveesta kuolleeksi
  par_terve95[i, 3] <- p[i] + (1 - p[i] - q[i]) * p[i] +
      (1 - p[i] - q[i])^2 * p[i] + (1 - p[i] - q[i])^3 * p[i] +
      (1 - p[i] - q[i])^4 * p[i] + (1 - p[i] - q[i])^5 * p[i] +
    q[i] * r[i] + (1 - p[i] - q[i]) * q[i] * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * r[i] +
      (1 - p[i] - q[i])^3 * q[i] * r[i] +
      (1 - p[i] - q[i])^4 * q[i] * r[i] +
    q[i] * (1 - r[i]) * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i]) * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i]) * r[i] +
      (1 - p[i] - q[i])^3 * q[i] * (1 - r[i]) * r[i] +
    q[i] * (1 - r[i])^2 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^2 * r[i] +
```

```
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^2 * r[i] +
    q[i] * (1 - r[i])^3 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^3 * r[i] +
    q[i] * (1 - r[i])^4 * r[i]

    # Vuosittaiset todennakoisyydet p, q ja r
    q[i] <- q_[i] / (1 + p_[i] + q_[i])
    r[i] <- r_[i] / (1 + r_[i])
    p[i] <- p_[i] / (1 + p_[i] + q_[i])

    # Regressiomallit
    log(q_[i]) <- inprod(beta_q, X[i, ])
    log(r_[i]) <- inprod(beta_r, X[i, ])
    log(p_[i]) <- inprod(beta_p, X[i, ])
}

## siirtyma diagnosoiduista 95 -> 01
for(i in (N85 + N_terve95 + 1):(N85 + N_terve95 + N_diag95)) {
  # Vasteet Bernoulli-jakaumasta parametrilla r95
  y[i] ~ dbern(r95[i])

  # Siirtyma diagnosoidusta kuolleeksi
  r95[i] <- r[i] + (1 - r[i]) * r[i] +
    (1 - r[i])^2 * r[i] + (1 - r[i])^3 * r[i] +
    (1 - r[i])^4 * r[i] + (1 - r[i])^5 * r[i]

  # Vuosittainen todennakoisyys r ja sen malli
  logit(r[i]) <- inprod(beta_r, X[i, ])
}

## Siirtyma terveista 01 -> 08
for(i in (N85 + N95 + 1):(N85 + N95 + N_terve01)) {
  # Vasteet kategorisesta jakaumasta parametrilla par_terve01
  y[i] ~ dcat(par_terve01[i, ])

  # Siirtyma terveesta terveeksi
  par_terve01[i, 1] <- (1 - p[i] - q[i])^7

  # Siirtyma terveesta diagnosoiduksi
  par_terve01[i, 2] <- q[i] * (1 - r[i])^6 +
    (1 - p[i] - q[i]) * q[i] * (1 - r[i])^5 +
    (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^4 +
    (1 - p[i] - q[i])^3 * q[i] * (1 - r[i])^3 +
    (1 - p[i] - q[i])^4 * q[i] * (1 - r[i])^2 +
```

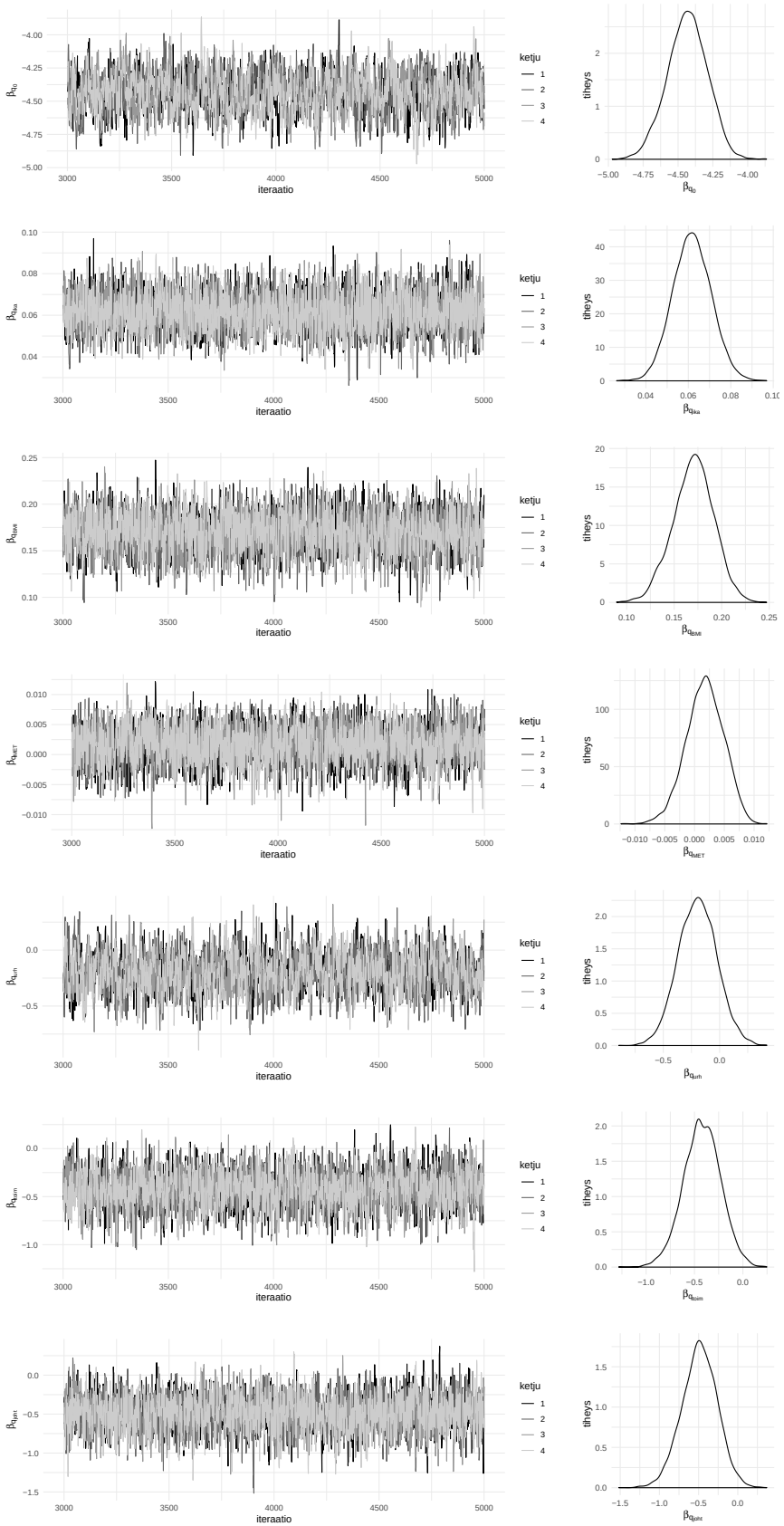
```
    (1 - p[i] - q[i])^5 * q[i] * (1 - r[i]) +
    (1 - p[i] - q[i])^6 * q[i]

  # Siirtyma terveesta kuolleeksi
  par_terve01[i, 3] <- p[i] + (1 - p[i] - q[i]) * p[i] +
      (1 - p[i] - q[i])^2 * p[i] + (1 - p[i] - q[i])^3 * p[i] +
      (1 - p[i] - q[i])^4 * p[i] + (1 - p[i] - q[i])^5 * p[i] +
      (1 - p[i] - q[i])^6 * p[i] +
    q[i] * r[i] + (1 - p[i] - q[i]) * q[i] * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * r[i] +
      (1 - p[i] - q[i])^3 * q[i] * r[i] +
      (1 - p[i] - q[i])^4 * q[i] * r[i] +
      (1 - p[i] - q[i])^5 * q[i] * r[i] +
    q[i] * (1 - r[i]) * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i]) * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i]) * r[i] +
      (1 - p[i] - q[i])^3 * q[i] * (1 - r[i]) * r[i] +
      (1 - p[i] - q[i])^4 * q[i] * (1 - r[i]) * r[i] +
    q[i] * (1 - r[i])^2 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^2 * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^2 * r[i] +
      (1 - p[i] - q[i])^3 * q[i] * (1 - r[i])^2 * r[i] +
    q[i] * (1 - r[i])^3 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^3 * r[i] +
      (1 - p[i] - q[i])^2 * q[i] * (1 - r[i])^3 * r[i] +
    q[i] * (1 - r[i])^4 * r[i] +
      (1 - p[i] - q[i]) * q[i] * (1 - r[i])^4 * r[i] +
    q[i] * (1 - r[i])^5 * r[i]

    # Vuosittaiset todennakoisyydet p, q ja r
    q[i] <- q_[i] / (1 + p_[i] + q_[i])
    r[i] <- r_[i] / (1 + r_[i])
    p[i] <- p_[i] / (1 + p_[i] + q_[i])

    # Regressiomallit
    log(q_[i]) <- inprod(beta_q, X[i, ])
    log(r_[i]) <- inprod(beta_r, X[i, ])
    log(p_[i]) <- inprod(beta_p, X[i, ])
}

## Siirtyma diagnosoiduista 01 -> 08
for(i in (N85 + N95 + N_terve01 + 1):(N85 + N95 + N_terve01 +
N_diag01)) {
  # Vasteet Bernoulli-jakaumasta parametrilla r01
```

```
    y[i] ~ dbern(r01[i])

    # Siirtyma diagnosoidusta kuolleeksi
    r01[i] <- r[i] + (1 - r[i]) * r[i] + (1 - r[i])^2 * r[i] +
      (1 - r[i])^3 * r[i] + (1 - r[i])^4 * r[i] +
      (1 - r[i])^5 * r[i] + (1 - r[i])^6 * r[i]

    # Vuosittainen todennakoisyys r ja sen malli
    logit(r[i]) <- inprod(beta_r, X[i, ])
  }

  # Priorit regressiokertoimille
  for(j in 1:n_p){
    # terve -> kuollut
    beta_p[j] ~ dnorm(0, 1 / 2)

    # terve -> diagnosoitu
    beta_q[j] ~ dnorm(0, 1 / 2)

    # diagnosoitu -> kuollut
    beta_r[j] ~ dnorm(0, 1 / 2)
  }

  # Priorit puuttuville kovariaateille
   for(n in 1:N){
    X[n, 3] ~ dnorm(0, 1 / 10)  # BMI
    X[n, 4] ~ dnorm(0, 1 / 700) # MET
    X[n, 6] ~ dbern(0.35)       # toimistotyontekija
    X[n, 7] ~ dbern(0.25)       # johtaja
    }
}"

# Mallin sovitus
# Datana vasteet y, kovariaattien matriisi X,
# matriisin X sarakkeiden lukumaara n_p, vuosittaiset
# terveiden lukumaarat N_terve85, N_terve95, N_terve01,
# vuosittaiset diagnosoitujen lukumaarat N_diag85, N_diag95,
# N_diag01 ja vuosittaiset havaintojen maarat N85, N95, N85.

model <- jags.model(textConnection(malli), data = par,
  n.chains = 4)
update(model, n.iter = 2000)
samp <- coda.samples(model, variable.names = c("beta_p", "beta_q",
  "beta_r"), n.iter = 2000)
```

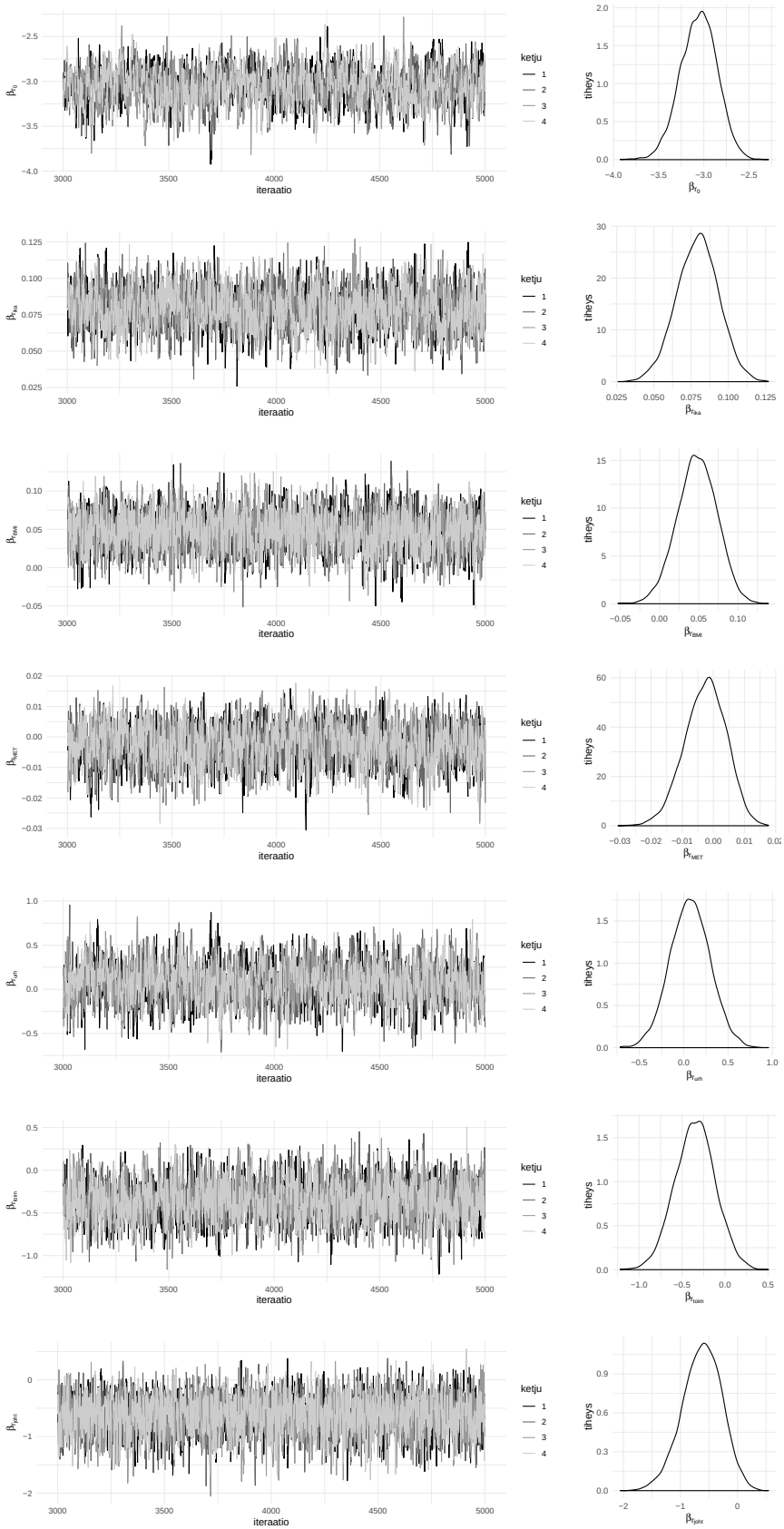
Liite 4. Mallin 2 sovitukseen liittyvät kuvat



Kuva 1: *Mallin 2 siirtymätodennäköisyyteen p liittyvien regressiokertoimien ketjut vasemmalla ja tiheyskuvaajat ketjuista lasketuille kertoimien estimaateille oikealla.*



Kuva 2: *Mallin 2 siirtymätodennäköisyyteen q liittyvien regressiokertoimien ketjut vasemmalla ja tiheyskuvaajat ketjuista lasketuille kertoimien estimaateille oikealla.*



Kuva 3: *Mallin 2 siirtymätodennäköisyyteen r liittyvien regressiokertoimien ketjut vasemmalla ja tiheyskuvaajat ketjuista lasketuille kertoimien estimaateille oikealla.*