



**TURUN
YLIOPISTO**

Mediaatioanalyysi tapahtuma-aikavasteen ja kategorisen
mediaattorin tilanteessa: tyypin 1 neurofibromatoosin suora ja
epäsuora vaikutus sydän- ja verisuonitautisairastavuuteen

Mikko Valtanen

Pro gradu -tutkielma
Helmikuu 2021

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck-järjestelmällä

TURUN YLIOPISTO

Matematiikan ja tilastotieteen laitos

VALTANEN, MIKKO: Mediaatioanalyysi tapahtuma-aikavasteen ja kategorisen mediaattorin tilanteessa: tyypin 1 neurofibromatoosin suora ja epäsuora vaikutus sydän- ja verisuonitautisairastavuuteen

Pro gradu -tutkielma, 38 s.

Tilastotiede

Helmikuu 2021

Mediaatioanalyysin tutkimuksessa on tapahtunut viime vuosikymmeninä merkittävää edistystä kontrafaktuaalisen kehikon omaksumisen myötä ja mediaatioanalyysiä koskeva kirjallisuus on laajentunut huomattavaa vauhtia. Tässä tutkielmassa esitellään modernin mediaatioanalyysin teoriaa. Mediaatiomallissa tarkasteltaville ns. luonnollisille kausaalivaikutuksille esitetään kontrolloitavan muuttujan perustasosta riippuvat määritelmät, näytetään miten ne suhteutuvat kirjallisuudessa usein käytettyihin määritelmiin ja selvitetään määritelmässä tehtävien valintojen vaikutusta tulosten tulkintaan. Luonnollisten vaikutusten lausekkeille esitellään sekä parametrittomat että niiden pohjalta johdettavat parametriset esitykset tietyin mallioletuksin. Uutena tuloksena tutkielmassa esitellään luonnollisille vaikutuksille parametriset lausekkeet diskreetin mediaattorin ja tapahtuma-aikavasteen tilanteessa, kun vasteelle oletetaan suhteellisten hasardien malli ja mediaattorille multinomimalli.

Mediaatioanalyysin sovelluksena tarkastellaan, miten tyypin 1 neurofibromatoosin (NF1) kausaalivaikutus iskeemiseen sydänsairauteen (IHD) jakautuu suoraan vaikutukseen ja kolmiluokkaisen sosioekonomisen statuksen (SES) kautta välittyvään epäsuoraan vaikutukseen. Asetelmassa huomioitavia sekoitavia tekijöitä ovat ensimmäisen asteen sukulaisen NF1-status sekä yksilön synnyinkunnan maaseutu-kaupunki-luokitus. Tulokset viittaavat, että SES:in kautta välittyvä epäsuora vaikutus on olemassa, mutta se on vähäinen verrattuna kaikista muista vaikutusmekanismeista koostuvaan suoraan vaikutukseen.

Asiasanat: mediaatioanalyysi, mediaattori, suora ja epäsuora vaikutus, kausaalivaikutus, tyypin 1 neurofibromatoosi.

Sisältö

1	Johdanto	1
2	Tyypin 1 neurofibromatoosi	2
3	Luonnolliset kausaalivaikutukset havainnoivassa tutkimusasetelmassa	4
3.1	Käsitteiden määrittely	4
3.1.1	<i>Do</i> -operaatio, kausaalivaikutus ja takaovipolku	4
3.1.2	Kontrafaktuaali	6
3.1.3	Luonnolliset vaikutukset	7
3.2	Perustasojen valinta	9
3.3	Luonnollisten vaikutusten identifioituvuus	10
4	Mediaatiomalli	13
4.1	Parametriton esitys	13
4.2	Parametriset esitykset lineaarisilla mallioletuksilla	15
4.2.1	Vaikutusten parametriset esitykset	16
4.2.2	Totaalivaikutuksen yhteys suoraan ja epäsuoraan vaikutukseen	17
4.2.3	Vaikutusten tulkinta	18
4.3	Vaikutukset suhteellisten hasardien mallissa	19
4.3.1	Vaikutusten määrittely	20
4.3.2	Tulkinta	22
5	Aineisto	22
6	Simulointitutkimus	24
6.1	Muuttujat	24
6.2	Simuloinnin tulokset	28
7	Aineiston analysointi	28
7.1	Mallioletukset	29
7.2	Tulokset	30
8	Pohdinta	34

1 Johdanto

Mediaatioanalyysiä sovelletaan tilanteissa, joissa kiinnostuksen kohteena olevalla selittäjällä — usein jollakin *altisteella* — on useita mahdollisia vaikutusmekanismeja vasteeseen ja tavoitteena on selvittää, miten altisteen kausaalivaikutus jakautuu eri mekanismien kesken. Tavallisesti vaikutukset jaetaan altisteen *suoraan vaikutukseen* ja *mediaattorien* kautta välittyviin *epäsuoriin vaikutuksiin*. Tietyn mediaattorin kautta kulkeva epäsuora vaikutus mittaa, kuinka suuri kausaalivaikutus on peräisin mekanismista, jossa altiste vaikuttaa kyseiseen mediaattoriin ja mediaattori vuorostaan vasteeseen. Suora vaikutus mittaa kausaalivaikutusta, joka ei välity minkään asetelmassa tarkasteltavan mediaattorin kautta. Mediaatioanalyysi erittelee kunkin mekanismin kautta välittyvän vaikutuksen suuruuden, minkä perusteella voidaan ymmärtää tutkittavan kausaalisen suhteen vaikutusmekanismeja ja arvioida potentiaalisten interventioiden merkitystä. [1–3]

Mediaatiomallit on perinteisesti määritelty lineaaristen rakenneyhtälömallien avulla, jolloin suora ja epäsuorat vaikutukset on johdettu sovitettujen lineaaristen mallien parametreista niin sanotuilla tulo- tai erotusmetodeilla [4–7]. Erotusmetodissa epäsuoran vaikutuksen tulkitaan olevan altisteen regressiokertoimen muutos, kun vasteelle sovitettuun malliin lisätään mediaattori selittäjäksi, ja suoran vaikutuksen olevan jäljelle jäävä altisteen regressiokerroin. Tulometodissa sovitetaan erilliset regressiomallit mediaattorille ja vasteelle, jolloin epäsuoran vaikutuksen tulkitaan olevan altisteen regressiokerroin mediaattorille sovitetussa mallissa kerrottuna mediaattorin regressiokertoimella vasteelle sovitetussa mallissa. Nämä menetelmät ovat kuitenkin sangen rajoituneita, sillä ne toimivat ainoastaan lineaaristen mallien tapauksessa ja olettamalla, että altisteen ja mediaattorin yhdysvaikutusta ei ole. Lisäksi vaikutusten määrittely mallien regressiokertoimien kautta aiheuttaa tulkinallisen ongelman, sillä tällöin määritelmät seuraavat mallin parametrisointia koskevista oletuksista, jolloin myös vaikutusten semanttinen sisältö on sidottu kyseisiin mallioletuksiin.

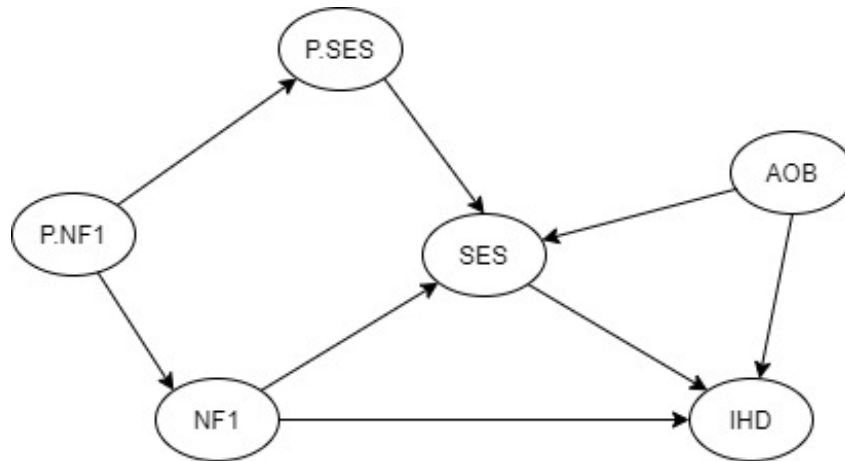
Kontrafaktuaalisen kausaliteetin käsitteen avulla mediaatioanalyysissä tarkasteltaville kausaalivaikutuksille on voitu esittää yleistetyt, tulkinallisesti mielekkäät määritelmät, jotka eivät edellytä oletuksia parametrisista malleista [8–11]. Vaikutusten parametrittomat määritelmät mahdollistavat mediaatioanalyysin soveltamisen myös epälineaarisisa tilanteissa. Mediaatiomalli voidaan määrittellä ilmiön taustalle oletettavaa kausaalirakennetta kuvaavan graa-

fin avulla, ja vaikutusten identifioituvuus perustuu altisteen, mediaattorin ja vasteen välisiin ehdollisiin riippumattomuusoletuksiin [12, 13], joiden toteutumista voidaan arvioida mediaatiomallia kuvaavan kausaaligraafin perusteella. Parametrittoman mediaatiomallin lausekkeiden kautta luonnollisille vaikutuksille voidaan edelleen johtaa parametriset esitykset, jotka määräytyvät mediaatiomallin parametrittoman muotoilun sekä mediaattorille ja vasteelle oletettavien parametrusten mallien perusteella. Vaikutusten määritelmät perustuvat kuitenkin yhä kontrafaktuaaleihin, eikä niiden tulkinta tällöin edelleenkään riipu oletettavien mallien parametrisista muodoista.

Tässä tutkielmassa sovelletaan mediaatioanalyysiä tavoitteena selvittää, miten tyypin 1 neurofibromatoosin vaikutus iskeemisiin sydänsairauksiin jakautuu suoraan ja epäsuoraan vaikutukseen. Tarkasteltava epäsuora vaikutus on tässä tapauksessa sosioekonomisen statuksen kautta välittyvä vaikutus. Tyypin 1 neurofibromatoosi ja sen tunnettuja yhteyksiä muihin sairauksiin, sekä ilmiön taustalle oletettava kausaalirakenne esitellään lyhyesti luvussa 2. Luvussa 3 esitellään kausaalipäättelyn peruskäsitteitä ja määritellään mediaatioanalyysin mukaiset niin sanotut luonnolliset kausaalivaikutukset sekä ehdot niiden identifioitumiselle. Luvussa 4 määritellään parametrin mediaatiomalli ja näytetään, miten siitä voidaan johtaa tulkittavissa oleva parametrinen esitysmuoto. Luvun tärkeimpänä tuloksena johdetaan suoralle ja epäsuoralle vaikutukselle tapahtuma-aikavasteen ja kategorisen mediaattorin tilanteessa parametriset lausekkeet, joita kirjallisuudessa ei ole aiemmin esitetty. Luvuissa 5, 6 ja 7 esitellään sovelluksessa käytetty aineisto, arvioidaan johdetun menetelmän tomivuutta sovelluksen mukaisessa tilanteessa simulointikokeella ja esitellään sovelluksen tulokset. Luvussa 8 pohditaan tutkielman tuloksia ja johdetun menetelmän haasteita.

2 Tyypin 1 neurofibromatoosi

Tyypin 1 neurofibromatoosi (NF1) on vallitsevasti periytyvä monielinoireyhtymä. Se aiheutuu mutaatiosta kromosomissa 17 sijaitsevassa *NF1*-geenissä, joka koodaa kasvurajoiteproteiini neurofibromiinia. NF1:n näkyvimpiä oireita ovat iholla näkyvät hyvänlaatuiset ääreishermoston kasvaimet, neurofibroomat, ja ihon maitokahviläiskät [14]. Vaikka ihon neurofibroomat eivät koskaan muutu pahanlaatuisiksi [15], NF1 on yhdistetty useisiin syöpiin, joista tavallisimpia ovat aivojen ja hermoston syövät. NF1-potilaiden syöpäilmaantuvuus on



Kuva 1: Tutkimusongelmaa kuvaava kausaaligraafi. Muuttujat NF1 ja SES ovat yksilön NF1-status ja yksilön sosioekonominen status. Muuttujat P.NF1 ja P.SES ovat ensimmäisen asteen sukulaisen NF1-status ja vanhempien SES, IHD iskeeminen sydänsairaus ja AOB yksilön synnyinkunnan maaseutu-kaupunki-luokitus.

muuhun väestöön verrattuna erityisen korkea nuorella iällä mutta tasaantuu noin 70 ikävuoteen mennessä muun väestön tasolle [15]. Syöpäilmaantuvuuden lisäksi NF1 on yhdistetty muun muassa kognitiivisiin vaikeuksiin sekä luuston epämuodostumiin. Sen sijaan diabeteksen esiintyvyyden NF1-potilaiden joukossa on todettu olevan muuta väestöä vähäisempi [16]. NF1:n vallitsevuuden Suomessa on arvioitu olevan noin 1/3000–1/2000 ja ilmaantuvuuden noin 1/2000 syntynyttä lasta kohti [17]. Puolet uusista NF1-tapauksista aiheutuu peritystä ja puolet uudesta mutaatiosta.

Tässä tutkielmassa selvitetään NF1:n kausaalista vaikutusta iskeemisiin sydänsairauksiin (IHD). Erityisen kiinnostuksen kohteena on, miten NF1:n vaikutus jakautuu toisaalta suoraan vaikutukseen ja toisaalta sosioekonomisen statuksen (SES) kautta välittyvään epäsuoraan vaikutukseen. NF1:een yhdistettyjen kognitiivisten ongelmien perusteella sen oletetaan vaikuttavan SESiin, ja SES puolestaan on yhdistetty merkittäviin sydänsairauksille altistaviin elintapatekijöihin, kuten tupakointiin, huonoon ruokavalioon ja vähäiseen liikuntaan. Mahdollisia NF1:n suoran vaikutuksen mekanismeja ovat poikkeavat tulehdusreaktiot, kohonneen syöpäilmaantuvuuden myötä syöpähoitojen aiheuttamat verenkiertoelimistön muutokset, ja toisaalta diabetesriskin aleneminen, mikä puolestaan laskisi sydänsairauksien riskiä. Lisäksi NF1:n tiedetään vaikuttavan erityisesti hermostopienasta kehittyviin kudoksiin, kuten sydänvalti-

moihin. Kuva 1 esittää ilmiön taustalle tässä työssä oletetun kausaalirakenteen. Polku $NF1 \rightarrow SES \rightarrow IHD$ kuvaa sosioekonomisen statuksen kautta välittyvää epäsuoraa vaikutusta, ja polku $NF1 \rightarrow IHD$ kuvaa $NF1$:n suoraa vaikutusta. Mahdollisina sekoittavina tekijöinä on otettu huomioon synnyinkunta (AOB) ja ensimmäisen asteen sukulaisen $NF1$ -status (P.NF1), jonka sekoittava vaikutus välittyy vanhempien sosioekonomisen statuksen (P.SES) kautta. Ensimmäisen asteen sukulaisen $NF1$ -status on ainoa tunnettu $NF1$:lle altistava tekijä, eikä ole syytä olettaa, että $NF1$ -geenin ja sydänsairauksille altistavien geenien välillä olisi yhteyttä, joten voidaan olettaa, että $NF1$:n ja IHD :n välillä ei ole sekoittavia tekijöitä.

3 Luonnolliset kausaalivaikutukset havainnoivassa tutkimusasetelmassa

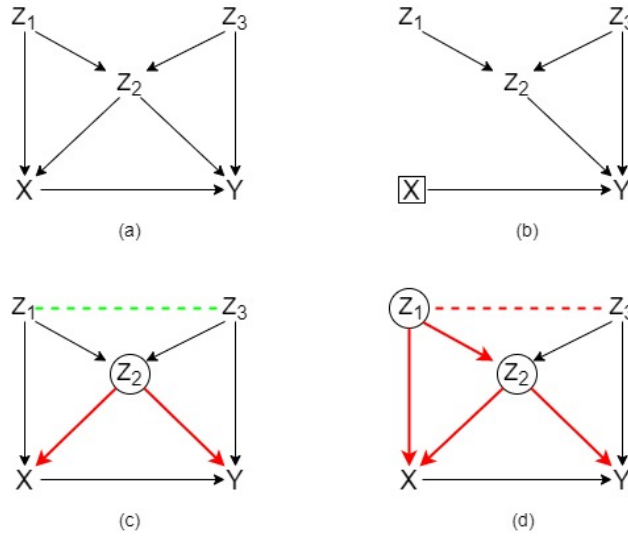
Kausaalivaikutuksista on eroteltavissa niin sanotut *luonnolliset vaikutukset* ja *kontrolloidut vaikutukset*. Mediaatioanalyysissä kontrolloidut vaikutukset ovat kiinnostuksen kohteina, kun halutaan tutkia jonkin intervention vaikutusta tilanteessa, jossa sekä altiste että mediaattori ovat kontrolloitavissa. Luonnolliset vaikutukset puolestaan ovat yleensä kiinnostavia tilanteissa, joissa mediaattori ei ole kontrolloitavissa.

Luonnollisten vaikutusten määrittely perustuu kontrafaktuaalisiin vasteisiin, joissa altisteelle asetetaan jokin arvo, ja mediaattorille annetaan arvo, jonka se luonnollisesti saisi, jos altisteella olisi jokin tietty, ei välttämättä sama kuin ensimmäisessä vaiheessa asetettu arvo. Nämä vasteet eivät yleisesti ottaen ole arvioitavissa kokeellisen aineiston perusteella, vaikka altiste ja mediaattori voitaisiin satunnaistaa [13]. Koska tämän tutkielman sovelluksessa kiinnostus on luonnollisissa vaikutuksissa, oletetaan jatkossa puhuttavan ainoastaan havainnoivista tutkimuksista ja vaikutuksilla tarkoitettavan luonnollisia vaikutuksia.

3.1 Käsitteiden määrittely

3.1.1 *Do*-operaatio, kausaalivaikutus ja takaovipolku

Kausaalinen vaikutus voidaan määritellä *do*-operaatioiden avulla. Näin ollen vaikka *do*-operaatioista puhutaankin yleensä kokeellisen asetelman yhteydessä, ne ovat käsitteellisesti olennainen osa kausaalipäätelyä myös havainnoivassa



Kuva 2: Graafit (a)–(d) kuvaavat *do*-operaation ja ehdollistamisen vaikutuksia kausaalisiin rakenteisiin. Neliö tarkoittaa, että muuttujaan on kohdistettu *do*-operaatio, ja ympyröinti tarkoittaa, että muuttujan suhteen on ehdollistettu. Vihreä (punainen) katkoviiva merkitsee avointa (vakioitua) suuntaamatonta yhteyttä ja punaiset nuolet vakioituja vaikutuksia.

asetelmassa. Kokeellisessa asetelmassa muuttujiin kohdistetaan interventioita eli vaikutetaan tutkittavaan tilanteeseen asettamalla halutuille muuttujille halutut arvot. Interventio voidaan kirjoittaa *do*-operaation avulla siten, että merkintä $do(X = x)$ vastaa interventiota, jossa muuttuja X asetetaan arvoon x . Interventio $do(X = x)$ tarkoittaa myös, että X on tämän jälkeen riippumaton kaikista siihen vaikuttavista muuttujista, jolloin vasteen Y muutos intervention $do(X = x)$ seurauksena on suoraan tulkittavissa muuttujan X kausaalivaikutukseksi vasteeseen Y . Näin ollen muuttujan X muutosta arvosta x_0 arvoon x_1 vastaavaa kausaalivaikutusta vasteeseen Y kuvaa erotus $E[Y|do(X = x_1)] - E[Y|do(X = x_0)]$. [18]

Havainnoivassa tutkimusasetelmassa ei voida toteuttaa *do*-operaatioita, jolloin havaitun muuttujan X yhteys vasteeseen Y saattaa olla peräisin sekoittavien tekijöiden vaikutuksesta. Sekoittava tekijä tarkoittaa yleisesti muuttujaa, josta kulkee kausaalinen polku sekä selittäjään että vasteeseen. Kausaalinen polku muuttujasta X muuttujaan Y tarkoittaa polkua, jossa jokainen muuttujasta X muuttujaan Y kulkevan polun nuoli kulkee kohti muuttujaa Y . *Do*-operaatioiden mukaiset kausaaliset vaikutukset voidaan estimoida havainnoivasta aineistosta, jos *do*-operaatioita vastaava sekoittajista riippumaton kausaalinen rakenne voidaan muodostaa ehdollistamalla taustamuuttujien

suhteen. Graafisesti tulkittuna tämä tarkoittaa, että niin sanotut *takaovipolut* muuttujasta X vasteeseen Y suljetaan [18]. Takaovipolku muuttujien X ja Y välillä tarkoittaa sellaista yhtenäistä polkua, jonka uloimmat nuolet osoittavat muuttujiin X ja Y . Jos muuttujien X ja Y välillä on avoin takaovipolku, niiden välinen vaikutus on sekoittunut, jolloin ko. kausaalivaikutusta ei voida estimoida harhattomasti. Muuttujien välisen vaikutuksen harhaton estimointi edellyttää siis kaikkien niiden välisten avoimien takaovipolkujen sulkemista.

Havainnoivassa asetelmassa takaovipolkujen sulkeminen tarkoittaa käytännössä, että ehdollistetaan jonkin sellaisen polkuun kuuluvan muuttujan suhteen, joka ei ole *törmäystekijä*. Törmäystekijä on muuttuja, johon osoittaa vähintään kaksi nuolta ja se itsessään sulkee takaovipolun. Muuttujat, joiden välillä on törmäystekijä, ovat riippumattomia, jos niiden välillä ei kulje muuta polkua, mutta ehdollistettaessa törmäystekijän suhteen siihen vaikuttavien muuttujien välille syntyy harhaanjohtava yhteys. Ehdollistaminen muuttujan suhteen yleisesti vakioi siitä lähtevät nuolet, jolloin kyseisen muuttujan syyt ja seuraukset ovat riippumattomia, ellei niiden välillä ole muuta polkua.

Tarkastellaan kausaalivaikutusten identifioituvuutta kuvan 2 avulla. Kuvassa 2(a) on esitetty muuttujien väliset kausaaliset suhteet. Muuttujasta X kulkevat vasteeseen Y takaovipolut $X \leftarrow Z_1 \rightarrow Z_2 \rightarrow Y$ ja $X \leftarrow Z_2 \rightarrow Y$. Kuvassa 2(b) muuttujaan X on kohdistettu interventio, jolloin siihen saapuvat nuolet häviävät, ja takaovipolut muuttujasta X vasteeseen Y sulkeutuvat. Kuvassa 2(c) on ehdollistettu kummallekin takaovipolulle yhteisen muuttujan Z_2 suhteen. Muuttuja Z_2 on kuitenkin törmäystekijä, joten sen suhteen ehdollistaminen avaa katkoviivalla merkityn yhteyden muuttujien Z_1 ja Z_3 välille, jolloin muuttujan X ja vasteen Y välille aukeaa uusi takaovipolku $X \leftarrow Z_1 - Z_3 \rightarrow Y$. Kaikkien takaovipolkujen sulkemiseksi on tällöin ehdollistettava lisäksi muuttujan Z_1 tai Z_3 suhteen, kuten kuvassa 2(d).

3.1.2 Kontrafaktuaali

Kontrafaktuaalilla tarkoitetaan vaihtoehtoista todellisuutta, jossa jokin havainnoituun vasteeseen mahdollisesti vaikuttava tekijä olisi ollut toisin, kuin se havaitussa todellisuudessa oli. Voidaan siis ajatella, että kun yksilöltä tehdään havainto tietyllä selittävän tekijän arvolla, samalla jäävät havaitsematta kontrafaktuaaliset vasteet kaikilla muilla selittävien tekijöiden mahdollisilla arvoilla. Kontrafaktuaalisia havaintoja ei siis koskaan havaita, mutta niiden jakaumia voidaan arvioida havainnoivasta aineistosta.

Kontrafaktuaalinen ajattelu perustuu tässä tutkielmassa tarkasteltavissa tilanteissa muuttujien T ja M vaihtoehtoisille arvoille. Käytetään tämän takia vasteelle Y merkintää $Y_{T_a M_{a^*}}$, kun altisteelle T on annettu arvo a , ja mediaattorille M on annettu se arvo, jonka se luonnollisesti saisi, jos altisteella olisi arvo a^* . Alaindeksit T_a ja M_{a^*} merkitsevät siis muuttujille T ja M annettuja arvoja, eli ne vastaavat *do*-operaatioita. Mikäli ehdollistetaan kovariaattien \mathbf{W} suhteen, käytetään merkintää $Y_{T_a M_{a^*}} | \mathbf{W}$. Kovariaateilla tarkoitetaan tässä yhteydessä mitä tahansa tilanteessa huomioon otettavia taustamuuttujia. Vasteet $Y_{T_a M_{a^*}}$ ja $Y_{T_{a^*} M_a}$ ovat aina kontrafaktuaalisia vasteita, sillä on mahdollista havaita vastetta, kun altisteella on jokin arvo, mutta mediaattorilla on arvo, joka sillä olisi ollut, jos altisteella olisikin ollut jokin toinen arvo.

3.1.3 Luonnolliset vaikutukset

Luonnollisten vaikutusten määrittely perustuu yksilötason kontrafaktuaalisiin vasteisiin. Koska yksilön kontrafaktuaalista vastetta ei koskaan tiedetä, luonnolliset vaikutukset voidaan laskea ainoastaan jonkin populaation tasolla. Tarkastellaan yleisesti muutosta $T : a^* \rightarrow a$, joka tarkoittaa, että altiste muuttuu tasolta a^* tasolle a . Oletetaan, että on olemassa kovariaattien joukko \mathbf{W} , jonka suhteen ehdollistamalla kausaalivaikutukset voidaan identifioida. Identifioituvuusehdot esitellään tarkemmin luvussa 3.3.

Totaalivaikutuksen (total effect), TE, määrittely on melko triviaali: se on vasteessa tapahtuva muutos, joka on seurausta altisteen muutoksesta. Totaalivaikutuksessa kontrafaktuaalinen ajattelu kohdistuu altisteeseen, ja mediaattorin sallitaan reagoivan altisteen muutokseen niin kuin se luonnollisesti reagoisikin. Yksilötasolla totaalivaikutus yksilölle i olisi

$$TE_i = Y_{T_a}^{(i)} - Y_{T_{a^*}}^{(i)} = Y_{T_a M_a}^{(i)} - Y_{T_{a^*} M_{a^*}}^{(i)}.$$

Vasteista kuitenkin vain toinen voidaan havaita, joten yksilötason totaalivaikutusta ei voida mitata, vaan sitä täytyy arvioida populaatiotasolla. Populaatiotasolla totaalivaikutus saadaan populaatioon kuuluvien yksilöiden vasteiden keskiarvona. Kovariaateille ehdollinen totaalivaikutus populaatiossa, jossa $\mathbf{W} = \mathbf{w}$, voidaan kirjoittaa

$$TE = E[Y_{T_a} | \mathbf{w}] - E[Y_{T_{a^*}} | \mathbf{w}] = E[Y_{T_a M_a} | \mathbf{w}] - E[Y_{T_{a^*} M_{a^*}} | \mathbf{w}], \quad (1)$$

jossa altisteen muutos vastaisi interventioiden *do*($T = a$) ja *do*($T = a^*$) toteuttamista, ja mediaattorin muutos on sen luonnollinen reaktio altisteen muutokseen.

Luonnollisen suoran vaikutuksen (*natural direct effect*), NDE , ja luonnollisen epäsuoran vaikutuksen (*natural indirect effect*), NIE , määrittelyt ovat tulkinnanvaraisempia, koska niissä joudutaan tekemään valinta kontrolloitavan muuttujan perustasosta. Perustasojen valintaan liittyy osittain altisteen mitta-asteikko. Tässä tutkielmassa keskitytään ensisijaisesti binääriseen altisteen tapaukseen, vaikka yleisesti altisteen mitta-asteikko voi olla mikä tahansa. Yleisesti olisi kuitenkin suotavaa, että altiste olisi binäärinen tai välimatka-asteikollinen.

Suora vaikutus on vasteessa tapahtuva muutos, kun $T : a^* \rightarrow a$, mutta mediaattori pysyy tasolla, joka sillä olisi, jos altiste olisi valitulla perustasolla $a^{(DE)}$. Tässä $a^{(DE)}$ on joko a^* tai a . Suora vaikutus vastaa kuvan 3(a) mukaisesta tilannesta, jossa nuoli $T \rightarrow M$ on katkaistu, mediaattorille asetettu arvoksi $M = M_{a^{(DE)}}$ ja tutkitaan, miten vaste muuttuu, kun altiste muuttuu tasolta a^* tasolle a . Altisteen ja mediaattorin välillä ei tällöin välity lainkaan informaatiota ja mediaattorin ja vasteen välinen informaatio pysyy vakiona. Suora vaikutus yksilölle i on siis

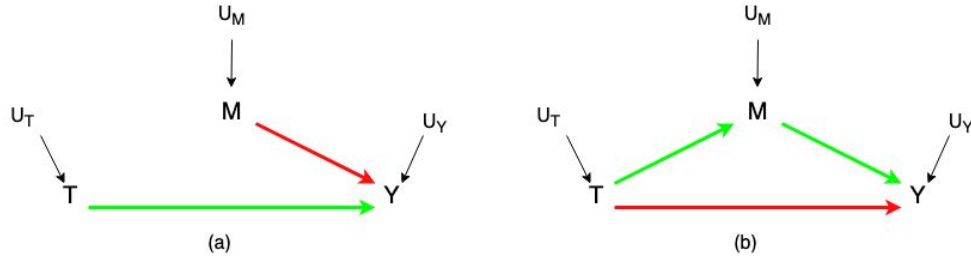
$$NDE_i = Y_{T_a M_{a^{(DE)}}}^{(i)} - Y_{T_{a^*} M_{a^{(DE)}}}^{(i)},$$

joka vastaisi mediaattoriin kohdistettua interventiota $do(M = M_{a^{(DE)}})$ ja altisteseen kohdistettuja interventioita $do(T = a)$ ja $do(T = a^*)$. On huomionarvoista, että mediaattoriin kohdistettu kontrafaktuaalinen interventio on täysin hypoteettinen tilanne, sillä sen kontrafaktuaalista arvoa ei tunneta, eikä kyseistä interventiota näin ollen voida toteuttaa edes kokeellisessa asetelmassa. Populaatiossa, jolla on kovariaattien arvo $\mathbf{W} = \mathbf{w}$, suora vaikutus voidaan kirjoittaa seuraavasti:

$$NDE = E[Y_{T_a M_{a^{(DE)}}} | \mathbf{w}] - E[Y_{T_{a^*} M_{a^{(DE)}}} | \mathbf{w}].$$

Epäsuora vaikutus määritellään vastaavasti vasteen muutoksena, kun mediaattori muuttuu tasolta M_{a^*} tasolle M_a , ja altiste pysyy perustasolla $a^{(IE)}$, jonka valitaan olevan joko a^* tai a . Tämä tilanne vastaa kuvan 3(b) mukaista tilannetta, jossa altisteen suora vaikutus on vakioitu valitun perustason mukaiseksi ja tarkastellaan vasteen muutosta, kun mediaattorin arvo muuttuu niin kuin se reagoisi altisteen muutokseen. Tässä tilanteessa altisteen muutoksen informaatio välittyy mediaattorin kautta vasteeseen, mutta suoran polun kautta kulkeva informaatio on vakioitu. Yksilölle i epäsuora vaikutus on tällöin

$$NIE_i = Y_{T_{a^{(IE)}} M_a}^{(i)} - Y_{T_{a^{(IE)}} M_{a^*}}^{(i)},$$



Kuva 3: Muuttujien välisen informaation kulku suorassa ja epäsuorassa vaikutuksessa. Kuva (a) esittää suoraa vaikutusta ja kuva (b) epäsuoraa vaikutusta. Punainen nuoli merkitsee, että polun kautta kulkeva vaikutus pidetään vakiona, eli sen kautta ei välity informaatiota. Vihreä nuoli merkitsee, että sen kautta kulkeva vaikutus on avoin muutoksille, eli se välittää informaatiota. Muuttujat U merkitsevät taustamuuttujia.

joka vastaisi mediaattoriin kohdistettuja interventioita $do(M = M_{a^*})$ sekä $do(M = M_a)$, ja altisteeseen kohdistettua interventiota $do(T = a^{(IE)})$. Kovariaattien $\mathbf{W} = \mathbf{w}$ määräämän populaation tasolla tämä vaikutus voidaan kirjoittaa

$$NIE = E[Y_{T_{a^{(IE)}}M_a} | \mathbf{w}] - E[Y_{T_{a^{(IE)}}M_{a^*}} | \mathbf{w}].$$

3.2 Perustasojen valinta

Luonnollisten vaikutusten tulkinta riippuu joissain tilanteissa niiden määrittelyssä tehtävistä perustasojen valinnoista. Kirjallisuudessa esitetyissä vaikutusten määritelmässä suoran vaikutuksen perustaso on yleensä altisteen muutoksen alkuarvo, mutta epäsuoran vaikutuksen perustaso vaihtelee lähteestä riippuen.

Pearl'in [11] esittämät vaikutusten määrittelyt vastaavat perustasojen valintaa siten, että $a^{(DE)} = a^{(IE)} = a^*$, kun $T : a^* \rightarrow a$. Toisin sanoen, perustasoiksi valittaisiin aina altisteen muutoksen alkuarvo, jolloin suora ja epäsuora vaikutus ovat

$$\begin{aligned} NDE &= E[Y_{T_a M_{a^*}} | \mathbf{w}] - E[Y_{T_{a^*} M_{a^*}} | \mathbf{w}], \\ NIE &= E[Y_{T_{a^*} M_a} | \mathbf{w}] - E[Y_{T_{a^*} M_{a^*}} | \mathbf{w}]. \end{aligned} \tag{2}$$

Jos totaali vaikutus (1) esitetään hajotelmana

$$TE = \left\{ E[Y_{T_a M_{a^*}} | \mathbf{w}] - E[Y_{T_{a^*} M_{a^*}} | \mathbf{w}] \right\} - \left\{ E[Y_{T_a M_{a^*}} | \mathbf{w}] - E[Y_{T_a M_a} | \mathbf{w}] \right\},$$

nähdään, että ensimmäinen termi vastaa Pearl'in määrittelemää suoraa vaikutusta, ja jälkimmäinen termi vastaa Pearl'in määritelmien mukaista negatii-

vista epäsuoraa vaikutusta, kun altiste muuttuisi vastakkaiseen suuntaan, eli $T : a \rightarrow a^*$. Täten Pearl'in mukaisilla määritelmillä pätee aina hajotelma

$$TE = NDE - NIE_r,$$

jossa NIE_r on epäsuora vaikutus käänteiseen suuntaan. Hajotelma voidaan kirjoittaa myös muodossa

$$TE = -\left\{ \mathbb{E}[Y_{T_a^* M_a} | \mathbf{w}] - \mathbb{E}[Y_{T_a M_a} | \mathbf{w}] \right\} + \left\{ \mathbb{E}[Y_{T_a^* M_a} | \mathbf{w}] - \mathbb{E}[Y_{T_a^* M_a^*} | \mathbf{w}] \right\}.$$

Tällöin totaalivaikutuksen lauseke voidaan lukea $TE = -NDE_r + NIE$, jossa NDE_r on puolestaan suora vaikutus käänteiseen suuntaan.

Vaihtoehtoiset vaikutusten määritelmät [19] seuraavat, kun valitaan $a^{(DE)} = a^*$ ja $a^{(IE)} = a$, kun $T : a^* \rightarrow a$, jolloin

$$\begin{aligned} NDE &= \mathbb{E}[Y_{T_a M_a^*} | \mathbf{w}] - \mathbb{E}[Y_{T_a^* M_a^*} | \mathbf{w}], \\ NIE &= \mathbb{E}[Y_{T_a M_a} | \mathbf{w}] - \mathbb{E}[Y_{T_a M_a^*} | \mathbf{w}]. \end{aligned} \tag{3}$$

Nämä määritelmät takaavat, että totaalivaikutukselle pätee aina hajotelma

$$TE = NDE + NIE.$$

Lisäksi nähdään, että vaihtoehtoisten määritelmien mukainen NIE on sama, kuin Pearl'in määritelmien mukainen $-NIE_r$.

Kun vaikutuksille oletetaan lineaariset mallit ja yhdysvaikutuksia ei ole, Pearl'in määritelmät ja vaihtoehtoiset määritelmät johtavat samaan tulokseen. Yhdysvaikutusten mukanaolo johtaa määritelmien välisiin tulkintaeroihin, joihin perehdytään tarkemmin kausaalivaikutusten parametrusten esitysten yhteydessä.

3.3 Luonnollisten vaikutusten identifioituvuus

Kausaalivaikutuksen identifioituvuudella tarkoitetaan tässä yhteydessä sitä, että ko. vaikutus voidaan estimoida harhattomasti ja tarkentuvasti. Luonnollisten vaikutusten identifioituminen edellyttää tiettyjen ehtojen toteutumista. Pearl [13] esittelee kolme vaihtoehtoista identifioituvuusehdot toteuttavaa oletusten joukkoa. Kutsutaan näitä Pearl'in mukaisesti joukoiksi A, B ja C. Joukko B sisältää Imain ym. [12] esittelemät kausaalivaikutusten identifiointiin riittävät riippumattomuusoletukset. Tässä luvussa esitellään Pearl'in artikkelin mukaiset ehtojen joukot A ja C, joiden sisältämät oletukset ovat Imain

ehtoja lievemmat. Ehtoja C käytetään luvussa 4, kun luonnollisille vaikutuksille johdetaan harhattomasti estimoituvat lausekkeet. Seuraavassa käytetään sanontaa ” \mathbf{W} puhdistaa M - Y -suhteen sekoittajista” (*deconfounds*), kun tarkoitetaan, että ehdollistamalla joukon \mathbf{W} suhteen, muuttujien M ja Y välillä ei ole sekoittavia tekijöitä, eli luvun 3.1.1 mukaisesti kaikki takaovipolut muuttujasta M muuttuunaan Y on suljettu.

Identifioituvuusehdot C

Oletetaan olevan olemassa sellaiset kolme mitattavissa olevaa kovariaattien joukkoa \mathbf{W}_1 , \mathbf{W}_2 ja \mathbf{W}_3 , että

- C1. altisteella ei ole kausaalivaikutusta mihinkään joukkoon \mathbf{W}_1 kuuluvaan muuttuunaan;
- C2. \mathbf{W}_1 puhdistaa M - Y -suhteen sekoittajista, kun altiste T pidetään vakiona;
- C3. joukko $\{\mathbf{W}_1, \mathbf{W}_2\}$ puhdistaa T - M -suhteen sekoittajista;
- C4. joukko $\{\mathbf{W}_1, \mathbf{W}_3\}$ puhdistaa $\{T, M\}$ - Y -suhteen sekoittajista.

Oletukset C1 ja C2 tarkoittavat graafisesti tulkittuna, että altisteesta ei kulje kausaalista polkua mihinkään muuttuunaan, jonka kontrollointi on välttämätöntä M - Y -suhteen puhdistamiseksi sekoittajista. Oletukset C3 ja C4 tarkoittavat, että niissä mainittujen suhteiden sekoittajat voidaan havaita ja niiden sekoittava vaikutus näin ollen kontrolloida.

Identifioituvuusehdot C rajoittuvat havainnoivaan tutkimusasetelmaan ja ovat siksi tämän tutkielman kannalta olennainen ehtojen joukko. Jos kuitenkin kokeellinen asetelma olisi mahdollinen, identifioituvuusehdot A tarjoavat lievemmat ehdot luonnollisten vaikutusten identifioimiseksi.

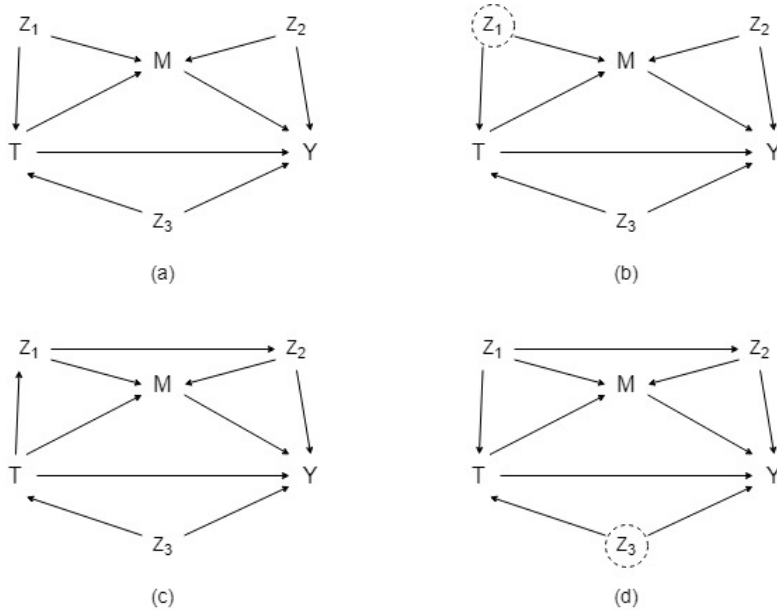
Identifioituvuusehdot A

Oletetaan olevan olemassa sellainen mitattavissa oleva kovariaattien joukko \mathbf{W} , että

- A1. altisteella ei ole kausaalivaikutusta mihinkään joukon \mathbf{W} muuttuunaan;
- A2. \mathbf{W} puhdistaa M - Y -suhteen sekoittajista, kun altiste T pidetään vakiona;
- A3. altisteen vaikutus mediaattoriin ehdolla $\mathbf{W} = \mathbf{w}$ on identifioitavissa;

A4. altisteen ja mediaattorin yhteinen vaikutus vasteeseen ehdolla $\mathbf{W} = \mathbf{w}$ on identifioitavissa.

Oletukset A1 ja A2 ovat samat kuin C1 ja C2, mutta A3 ja A4 laajentavat tarkastelun myös kokeelliseen asetelmaan. Vaikka suhteet T – Y ja $\{T, M\}$ – Y voidaan tällöin identifioida kokeellisesti, oletus A2 on kuitenkin aina toteutettava havainnoivassa asetelmassa. Tämä johtuu siitä, että M – Y -suhteen puhdistaminen sekoittajista tarkoittaisi kokeellisessa asetelmassa mediaattorin satunnaistamista, mikä ei kuitenkaan sallisi mediaattorin luonnollista reaktiota altisteen muutokseen.



Kuva 4: Graafit (a)–(d) esittävät erilaisia kausaalirakenteita. Katkoviivalla ympäröity muuttuja merkitsee havaitsematonta muuttujaa.

Tarkastellaan ehtoja kuvan 4 avulla. Havainnoivassa asetelmassa ainoastaan kuvan 4(a) tilanteessa luonnolliset vaikutukset ovat identifioitavissa. Muuttuja Z_2 toteuttaa oletukset C1 ja C2, koska se puhdistaa M – Y -suhteen sekoittajista, eikä siihen ole kausaalista polkua altisteesta. Lisäksi joukot $\{Z_1, Z_2\}$ ja $\{Z_2, Z_3\}$ toteuttavat oletukset C3 ja C4. Sen sijaan kuvan 4(b) tilanteessa muuttuja Z_1 on havaitsematon, jolloin oletus C3 ei täyty. Kuvassa 4(c) kaikki muuttujat ovat havaittavissa, mutta altisteesta kulkee kausaalinen polku M – Y -suhteen sekoittajaan Z_2 , mikä rikkoo oletuksen C1. Kuvan 4(d) tilanteessa jälleen muuttuja Z_3 on havaitsematon, joten oletus C4 ei täyty.

Mikäli kokeellinen asetelma olisi mahdollinen, myös kuvien 4(b) ja 4(d) tilanteissa luonnolliset vaikutukset olisivat identifioitavissa. Oletuksissa A3 ja

A4 mainittavat muuttujalle Z_2 ehdolliset suhteet voidaan tällöin identifoida kokeellisesti, ja myös oletukset A1 ja A2 täyttyvät, koska Z_2 voidaan havainnoida, eikä siihen ole kausaalista polkua altisteesta. Kuvan 4(c) tilanteessa oletus A1 ei kuitenkaan täyty, koska altisteesta on kausaalinen polku M - Y -suhteen sekoittajaan Z_2 .

4 Mediaatiomalli

Mediaatiomalli koostuu altisteen, mediaattorien, vasteen ja olennaisten taustamuuttujien määräämisestä sekä oletuksista näiden välisistä kausaalisista suhteista. Mediaatiomallin voidaan siis ajatella määräytyvän jo taustalle oletettavan kausaalirakenteen perusteella, ja mallin tarkoituksena on tuottaa arviot luonnollisille vaikutuksille. Luonnolliset vaikutukset on toistaiseksi esitetty ai-noastaan muodossa, joka vastaa *do*-operaatioiden suorittamista altisteelle ja mediaattorille. *Do*-operaatioiden suorittaminen ei kuitenkaan ole havainnoivas-sa asetelmassa mahdollista. Tässä luvussa näytetään, miten identifioituvuuseh-tojen toteutuessa luonnollisille vaikutuksille saadaan lausekkeet, jotka ovat ar-voitavissa havainnoivassa asetelmassa. Luonnollisten vaikutusten lausekkeet johdetaan ensin parametrittömästi, minkä jälkeen näytetään, miten niille voi-daan parametrusten mallioletusten myötä johtaa parametriset esitysmuodot. Parametrusten esitysten yhteydessä näytetään myös, miten epälineaarisisissa ti-lanteissa luonnollisten vaikutusten tulkinta riippuu perustasojen valinnoista.

4.1 Parametriton esitys

Luonnolliset vaikutukset määriteltiin populaatiotason odotusarvona kahden yksilötason vasteen erotuksesta. Vaikutuksia ei kuitenkaan ole mahdollista suo-raan mitata, koska ne perustuvat kontrafaktuaalisiin vasteisiin, joita ei voi-da havaita. Kontrafaktuaaliset vasteet estimoidaan ketjusääntöä käyttäen vas-teen ja mediaattorin ehdollisista jakaumista. Vaste $E[Y_{T_a M_{a^*}} | \mathbf{w}]$ vastaisi *do*-operaatioiden suorittamista altisteelle ja mediaattorille. Mediaattoriin kohdis-tettu *do*-operaatio on nyt kuitenkin käsitteellisestikin ongelmallinen, koska se ei populaatiotasolla määrää mediaattorille tiettyä arvoa, vaan ehdollisen ja-kauman.

Ketjusääntöä käyttäen suoran ja epäsuoran vaikutuksen kaavat populaa-tiolle, jolla on kovariaattien arvo $\mathbf{W} = \mathbf{w}$, voidaan kirjoittaa *do*-operaatioiden

avulla seuraavasti [11]:

$$\begin{aligned}
NDE &= \mathbb{E}[Y_{T_a M_a(D E)} | \mathbf{w}] - \mathbb{E}[Y_{T_{a^*} M_a(D E)} | \mathbf{w}] \\
&= \sum_m \left[\mathbb{E}[Y | do(T_a, m), \mathbf{w}] P_M(m | do(T_a(D E)), \mathbf{w}) - \right. \\
&\quad \left. \mathbb{E}[Y | do(T_{a^*}, m), \mathbf{w}] P_M(m | do(T_a(D E)), \mathbf{w}) \right] \\
&= \sum_m \left[\left(\mathbb{E}[Y | do(T_a, m), \mathbf{w}] - \mathbb{E}[Y | do(T_{a^*}, m), \mathbf{w}] \right) P_M(m | do(T_a(D E)), \mathbf{w}) \right],
\end{aligned}$$

$$\begin{aligned}
NIE &= \mathbb{E}[Y_{T_{a(IE)} M_a} | \mathbf{w}] - \mathbb{E}[Y_{T_{a(IE)} M_{a^*}} | \mathbf{w}] \\
&= \sum_m \left[\mathbb{E}[Y | do(T_{a(IE)}, m), \mathbf{w}] P_M(m | do(T_a), \mathbf{w}) - \right. \\
&\quad \left. \mathbb{E}[Y | do(T_{a(IE)}, m), \mathbf{w}] P_M(m | do(T_{a^*}), \mathbf{w}) \right] \\
&= \sum_m \left[\left(P_M(m | do(T_a), \mathbf{w}) - P_M(m | do(T_{a^*}), \mathbf{w}) \right) \mathbb{E}[Y | do(T_{a(IE)}, m), \mathbf{w}] \right].
\end{aligned}$$

Jotta luonnolliset vaikutukset voidaan estimoida havainnoivasta aineistosta, *do*-operaatiot on korvattava havaittavissa olevilla ehdollisilla jakaumilla. Mikäli kovariaattijoukko \mathbf{W} toteuttaa identifioituvuusehdot C, *do*-operaatiot voidaan korvata ehdollisilla todennäköisyyksillä [13] ja vaikutukset populaatiolle, jolla on kovariaattien arvo $\mathbf{W} = \mathbf{w}$, voidaan kirjoittaa seuraavasti:

$$\begin{aligned}
TE &= \mathbb{E}[Y | T_a, \mathbf{w}] - \mathbb{E}[Y | T_{a^*}, \mathbf{w}], \\
NDE &= \sum_m \left[\left(\mathbb{E}[Y | T_a, m, \mathbf{w}] - \mathbb{E}[Y | T_{a^*}, m, \mathbf{w}] \right) P_M(m | T_a(D E), \mathbf{w}) \right], \quad (4) \\
NIE &= \sum_m \left[\left(P_M(m | T_a, \mathbf{w}) - P_M(m | T_{a^*}, \mathbf{w}) \right) \mathbb{E}[Y | T_{a(IE)}, m, \mathbf{w}] \right].
\end{aligned}$$

Toistaiseksi vaikutuksia on tarkasteltu erikseen kovariaattien muodostamien ositteiden määräämissä osapopulaatioissa. Jos kovariaattien jakauma tunnetaan, vaikutukset voidaan yleistää isompiin populaatioihin käyttämällä ketjusääntöä. Tarkastellaan esimerkiksi kuvan 4(a) tilannetta, jossa muuttujajoukko $\{Z_1, Z_2, Z_3\}$ toteuttaa identifioituvuusoletukset. Esimerkiksi suora

vaikutus muuttujasta Z_3 riippumattomalle populaatiolle voidaan tällöin kirjoittaa muotoon:

$$\begin{aligned}
NDE &= \sum_{z_3} \sum_m \left[\left(\mathbb{E}[Y|T_a, m, z_1, z_2, z_3] - \mathbb{E}[Y|T_{a^*}, m, z_1, z_2, z_3] \right) \times \right. \\
&\quad \left. P_M(m|T_{a^{(DE)}}, z_1, z_2, z_3) P_{Z_3}(z_3|z_1, z_2) \right] \\
&= \sum_{z_3} \sum_m \left[\left(\mathbb{E}[Y|T_a, m, z_2, z_3] - \mathbb{E}[Y|T_{a^*}, m, z_2, z_3] \right) \times \right. \\
&\quad \left. P_M(m|T_{a^{(DE)}}, z_1, z_2) P_{Z_3}(z_3) \right],
\end{aligned}$$

jossa jälkimmäinen yhtäsuuruus seuraa muuttujien välisistä riippumattomuuksista.

Vastaavasti jos vaikutukset halutaan arvioida kaikista kovariaateista riippumattomalle populaatiolle ne summataan ketjusäännön mukaisesti yli kaikkien kovariaattien jakauman. Olettaen, että tarvittavat jakaumat tunnetaan, suoran ja epäsuoran vaikutuksen kaavoiksi saataisiin kuvan 4(a) mukaisessa tilanteessa

$$\begin{aligned}
NDE &= \sum_{z_2} \sum_m \left\{ \left(\sum_{z_3} \left(\mathbb{E}[Y|T_a, m, z_2, z_3] - \mathbb{E}[Y|T_{a^*}, m, z_2, z_3] \right) P_{Z_3}(z_3) \right) \times \right. \\
&\quad \left. \sum_{z_1} \left(P_M(m|T_{a^{(DE)}}, z_1, z_2) P_{Z_1}(z_1) \right) \right\} P_{Z_2}(z_2), \\
NIE &= \sum_{z_2} \sum_m \left\{ \left(\sum_{z_1} \left[P_M(m|T_a, z_1, z_2) - P_M(m|T_{a^*}, z_1, z_2) \right] P_{Z_1}(z_1) \right) \times \right. \\
&\quad \left. \sum_{z_3} \left(\mathbb{E}[Y|T_{a^{(IE)}}, m, z_2, z_3] P_{Z_3}(z_3) \right) \right\} P_{Z_2}(z_2).
\end{aligned}$$

4.2 Parametriset esitykset lineaarisilla mallioletuksilla

Jos mediaattorille ja vasteelle voidaan olettaa parametriset mallit, luonnollisille vaikutuksille voidaan johtaa parametriset esitykset. Luonnolliset vaikutukset määriteltiin jo yleisellä tasolla kahden vasteen odotusarvojen erotuksena, joten parametristen mallien oletaminen ei vaikuta luonnollisten vaikutusten määrittelyyn, eikä siten tulkintaankaan, vaan ainoastaan mahdollistaa luonnollisille vaikutuksille valittujen mallien mukaiset parametriset esitykset. Tämä tarkoittaa, että parametristen esitysten pohjalta tehtävät tulkinnot ovat

järkeviä ainoastaan silloin, kun oletetut mallit vastaavat hyvin ilmiön taustalla olevaa todellisuutta. Parametristen mallien oletaminen siis toisaalta tarjoaa mielekkään esitystavan, mutta toisaalta lisää riskiä päätyä virheellisiin päätelmiin.

Luvussa 3.2 mainittiin, että luonnollisten vaikutusten tulkinta on yksiselitteistä, kun muuttujien välillä ei ole yhdysvaikutuksia, mutta jos yhdysvaikutuksia on, luonnollisten vaikutusten tulkinta riippuu perustasojen valinnasta. Tämä käy hyvin ilmi, kun tarkastellaan lineaarisia malleja ilman yhdysvaikutusta ja yhdysvaikutuksen kanssa.

4.2.1 Vaikutusten parametriset esitykset

Tarkastellaan aluksi kuvan 4(a) mukaista tilannetta ja oletetaan vasteelle ja mediaattorille lineaariset mallit ilman yhdysvaikutuksia. Merkitään malleissa käytettäviä sekoittajien joukkoja $\mathbf{W}_1 = \{Z_1, Z_2\}$ ja $\mathbf{W}_2 = \{Z_2, Z_3\}$. Mediaattorin ja vasteen odotusarvot ovat

$$\begin{aligned} E[M|T = a, \mathbf{W}_1 = \mathbf{w}_1] &= \beta_0 + \beta_1 a + \boldsymbol{\beta}'_2 \mathbf{w}_1, \\ E[Y|T = a, M = m, \mathbf{W}_2 = \mathbf{w}_2] &= \theta_0 + \theta_1 a + \theta_2 m + \boldsymbol{\theta}'_3 \mathbf{w}_2. \end{aligned}$$

Luonnollisille vaikutuksille saadaan tällöin kontrafaktuaalisten määritelmien mukaiset parametriset esitystavat

$$\begin{aligned} TE &= \left[\theta_0 + \theta_1 a + \theta_2 (\beta_0 + \beta_1 a + \boldsymbol{\beta}'_2 \mathbf{w}_1) + \boldsymbol{\theta}'_3 \mathbf{w}_2 \right] - \\ &\quad \left[\theta_0 + \theta_1 a^* + \theta_2 (\beta_0 + \beta_1 a^* + \boldsymbol{\beta}'_2 \mathbf{w}_1) + \boldsymbol{\theta}'_3 \mathbf{w}_2 \right] \\ &= [\theta_1 + \theta_2 \beta_1] (a - a^*), \end{aligned}$$

$$\begin{aligned} NDE &= \left[\theta_0 + \theta_1 a + \theta_2 (\beta_0 + \beta_1 a^{(DE)} + \boldsymbol{\beta}'_2 \mathbf{w}_1) + \boldsymbol{\theta}'_3 \mathbf{w}_2 \right] - \\ &\quad \left[\theta_0 + \theta_1 a^* + \theta_2 (\beta_0 + \beta_1 a^{(DE)} + \boldsymbol{\beta}'_2 \mathbf{w}_1) + \boldsymbol{\theta}'_3 \mathbf{w}_2 \right] \\ &= \theta_1 (a - a^*), \end{aligned}$$

$$\begin{aligned} NIE &= \left[\theta_0 + \theta_1 a^{(IE)} + \theta_2 (\beta_0 + \beta_1 a + \boldsymbol{\beta}'_2 \mathbf{w}_1) + \boldsymbol{\theta}'_3 \mathbf{w}_2 \right] - \\ &\quad \left[\theta_0 + \theta_1 a^{(IE)} + \theta_2 (\beta_0 + \beta_1 a^* + \boldsymbol{\beta}'_2 \mathbf{w}_1) + \boldsymbol{\theta}'_3 \mathbf{w}_2 \right] \\ &= \theta_2 \beta_1 (a - a^*). \end{aligned}$$

Suora ja epäsuora vaikutus summautuvat perustasojen valinnoista riippumatta totaali-vaikutukseksi, eli aina pätee, että $TE = NDE + NIE$, ja oletettujen mallien mukainen parametrinen esitystapa määräytyy yksikäsitteisesti.

Lisätään sitten vasteen malliin altisteen ja mediaattorin yhdysvaikutus. Vasteen odotusarvoksi saadaan

$$E[Y|T = a, M = m, W_2 = w_2] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \boldsymbol{\theta}'_4 \mathbf{w}_2,$$

josta seuraa, että

$$\begin{aligned} TE &= [\theta_1 + \theta_2 \beta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 (a + a^*) + \theta_3 \boldsymbol{\beta}'_2 \mathbf{w}_1] (a - a^*), \\ NDE &= [\theta_1 + \theta_3 (\beta_0 + \beta_1 a^{(DE)} + \boldsymbol{\beta}'_2 \mathbf{w}_1)] (a - a^*), \\ NIE &= [\theta_2 \beta_1 + \theta_3 \beta_1 a^{(IE)}] (a - a^*). \end{aligned} \quad (5)$$

Kaavoista nähdään, miten perustasojen valinta vaikuttaa luonnollisiin vaikutuksiin ja miten vaikutusten parametriset lausekkeet eroavat luvussa 3.2 esiteltujen Pearl'in (2) sekä vaihtoehtoisten (3) määritelmien mukaisissa esityksissä. Lisäksi nähdään, että välimatka-asteikollisen altisteen tapauksessa vaikutusten suuruus ei välttämättä ole sama kaikilla samansuuruisilla altisteen muutoksilla.

4.2.2 Totaali-vaikutuksen yhteys suoraan ja epäsuoraan vaikutukseen

Kun vasteen mallissa on yhdysvaikutus, totaali-vaikutuksen yhteys suoraan ja epäsuoraan vaikutukseen riippuu siitä, mitä luonnollisten vaikutusten määrittelmää käytetään. Linearisessa tilanteessa suoran ja epäsuoran vaikutuksen summa on

$$NDE + NIE = [\theta_1 + \theta_2 \beta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 (a^{(DE)} + a^{(IE)}) + \theta_3 \boldsymbol{\beta}'_2 \mathbf{w}_1] (a - a^*),$$

josta nähdään, että perustasojen valinta vaikuttaa yhdysvaikutuksen parametrin sisältävään termiin, ja vaikutusten summa selvästikin riippuu perustasojen valinnoista.

Vähentämällä totaali-vaikutuksesta suora ja epäsuora vaikutus, saadaan

$$\begin{aligned} TE - (NDE + NIE) &= [\theta_1 + \theta_2 \beta_1 + \theta_3 (\beta_0 + \beta_1 (a + a^*) + \boldsymbol{\beta}'_2 \mathbf{w}_1)] (a - a^*) - \\ &\quad [\theta_1 + \theta_2 \beta_1 + \theta_3 (\beta_0 + \beta_1 (a^{(DE)} + a^{(IE)}) + \boldsymbol{\beta}'_2 \mathbf{w}_1)] (a - a^*) \\ &= \theta_3 \beta_1 [(a + a^*) - (a^{(DE)} + a^{(IE)})] (a - a^*). \end{aligned} \quad (6)$$

Hajotelma $TE = NDE + NIE$ on siis aina voimassa, jos $\theta_3 = 0$, eli jos yhdysvaikutusta ei ole. Kun mallissa on yhdysvaikutus, hajotelma pätee, jos valitut perustasot ovat altisteen muutoksen erisuuret tasot, eli $a^{(DE)} \neq a^{(IE)}$ ($a^{(DE)}$ ja $a^{(IE)} \in \{a^*, a\}$), mikä vastaa vaikutusten vaihtoehtoista määritelmää (3). Välimatka-asteikollisella altisteella hajotelman voidaan sanoa pätevän aina, kun $a^{(DE)} + a^{(IE)} = a^* + a$.

4.2.3 Vaikutusten tulkinta

Totaalivaikutukselle esiteltiin aiemmin hajotelma

$$TE = \left\{ E[Y_{T_a M_{a^*}} | \mathbf{w}] - E[Y_{T_{a^*} M_{a^*}} | \mathbf{w}] \right\} + \left\{ E[Y_{T_a M_a} | \mathbf{w}] - E[Y_{T_a M_{a^*}} | \mathbf{w}] \right\}.$$

Totaalivaikutus voidaan siis jakaa kahteen osaan, joista ensimmäisessä Y reagoi altisteen T muutokseen ja toisessa altisteen muutoksen seurauksena muuttuneen mediaattorin M muutokseen. Tällöin totaalivaikutus sisältää kummas-takin vaiheesta tulevan yhdysvaikutustermin: ensimmäinen osa sisältää yhdysvaikutustermin, jossa kontrolloitavan muuttujan perustasona on a^* , ja toinen osa sisältää yhdysvaikutustermin, jossa kontrolloitavan muuttujan perustasona on a . Tätä yhdysvaikutustermin muutosta kuvaa totaalivaikutukselle johdetun, lineaaristen mallien mukaisen parametrinen esityksen (5) termi

$$\theta_3 \beta_1 (a + a^*).$$

Aiemmin esiteltyjen luonnollisten vaikutusten määritelmien ero liittyy yhdysvaikutukseen. Pearl'in määritelmässä molemmat perustasot valitaan altisteen alkutilaksi a^* , jolloin lineaarisessa tilanteessa sijoittamalla valitut perustasot kaavaan (6), totaalivaikutukselle saadaan lauseke [20]

$$TE = NDE + NIE + \theta_3 \beta_1 (a - a^*)(a - a^*),$$

jossa viimeinen termi tulee siitä, että summa $NDE + NIE$ ei tällöin sisällä yhdysvaikutustermiä altisteen lopputasolla a . Jos toisen vaikutuksen perustasoksi valitaankin altisteen lopputaso a , NDE ja NIE summautuvat totaalivaikutukseksi, mutta toinen vaikutuksista pysyy silti ennallaan. Näin ollen voidaan päätellä, että jos vaikutuksen perustasoksi valitaan altisteen lopputaso, siihen sisällytetään myös termi $\theta_3 \beta_1 (a - a^*)$. Vaihtoehtoisessa määritelmässä (3) yhdysvaikutustermin muutos sisältyy siis kokonaan epäsuoraan vaikutukseen, kun taas Pearl'in määritelmässä yhdysvaikutustermi on kummassakin vaikutuksessa sama ja sen muutos jätetään erilliseksi osaksi kokonaisvaikutusta.

Pearl esittää kaksi eri tapaa arvioida epäsuoran vaikutuksen merkitystä tarkasteltavassa ilmiössä. Luvussa 3.2 esitellyillä Pearl’in mukaisilla määritelmillä totaalivaikutuksen hajotelman toteuttava $-NIE_r$ mittaa mediaation aiheuttamaa, ja NIE puolestaan mediaation selittämää vaikutusta. Mediaation selittämällä vaikutuksella tarkoitetaan sitä vaikutusta, jonka pelkkä mediaattorin muutos saisi aikaan ilman, että altiste muuttuu alkutasoltaan. Mediaation aiheuttamassa vaikutuksessa puolestaan altiste on mediaattorin muuttuessa jo lopputasollaan, jolloin mediaation aiheuttamaan vaikutukseen sisältyy myös yhdysvaikutuksen muutos. Se kuvaa siis kaikkea vaikutusta, joka olisi jäänyt tapahtumatta ilman mediaattorin muutosta. Mediaation selittämä vaikutus, NIE , ja aiheuttama vaikutus, $-NIE_r$, saadaan kaavoista

$$\begin{aligned} NIE &= E[Y_{T_{a^*}M_a}|\mathbf{w}] - E[Y_{T_{a^*}M_{a^*}}|\mathbf{w}], \\ -NIE_r &= TE - NDE, \end{aligned}$$

joissa NIE on siis määritelmänsä mukainen odotusarvojen erotus, ja $-NIE_r$ saadaan vähentämällä totaalivaikutuksesta suora vaikutus.

Myös suoralle vaikutukselle voidaan tehdä tulkinta sen aiheuttamasta ja selittämästä vaikutuksesta [20]. Sisällyttämällä yhdysvaikutuksen muutos suoraan vaikutukseen, saadaan suoran vaikutuksen aiheuttama osuus, joka vastaisi Pearl’in määritelmien mukaista negatiivista käänteistä suoraa vaikutusta, $-NDE_r$. Tämä saadaan vähentämällä totaalivaikutuksesta epäsuoran vaikutuksen selittämä osuus.

4.3 Vaikutukset suhteellisten hasardien mallissa

Kun vastemuuttuja on ’aika tapahtumaan’, kiinnostuksena on usein selvittää, kuinka eri altisteet vaikuttavat tapahtuman *hasardiin*. Hasardi tarkoittaa vauhtia, jolla yksilön todennäköisyys kohdata päätetapahtuma kasvaa ajanhetkellä t , sillä ehdolla, että yksilö ei ole vielä kohdannut päätetapahtumaa ajanhetkeen t mennessä. Myös luonnolliset vaikutukset voidaan esittää hasardiskaalalla [21]. Tässä luvussa johdetaan mediaatiomallin mukaiset vaikutusten lausekkeet hasardiskaalalla, kun hasardille oletetaan suhteellisten hasardien malli ja mediaattorille multinomimalli.

Suhteellisten hasardien mallissa oletetaan, että hasardien suhde kovariaatioiden määräämissä ositteissa on ajan suhteen vakio. Selittäjien oletetaan vaikuttavan hasardiin multiplikatiivisesti, jolloin selittäjille ehdollinen hasardi ajan-

hetkellä t on

$$\lambda(t|a, m, \mathbf{w}) = \lambda(t|0, 0, 0)e^{\gamma_1 a + \gamma_2 m + \gamma_3 am + \gamma_4' \mathbf{w}}.$$

Yllä $\lambda(t|0, 0, 0)$ on perustason hasardi. Kun hasardisuhteiden estimointi perustuu ns. Cox'in malliin, päättely perustuu osittaisuskottavuuteen, joka ei huomioi perustason hasardia ja se jää näin ollen estimoimatta [22]. Cox'in mallin perusteella pystytään näin ollen arvioimaan ainoastaan hasardiin suhdetta.

4.3.1 Vaikutusten määrittely

Kun vasteen hasardifunktiolle oletetaan multiplikatiivinen malli, myös luonnolliset vaikutukset on loogista määrittellä multiplikatiivisella skaalalla. Cox'in mallin tapauksessa tämä on käytännöllisestikin suotavaa, koska hasardiin erosta ei voida laskea tuntematta perustason hasardia. Vaikutukset ovat tällöin

$$\begin{aligned} TE^{(HR)} &= \lambda_{T_a} / \lambda_{T_{a^*}}, \\ NDE^{(HR)} &= \lambda_{T_a M_{a(DE)}} / \lambda_{T_{a^*} M_{a(DE)}}, \\ NIE^{(HR)} &= \lambda_{T_{a(IE)} M_a} / \lambda_{T_{a(IE)} M_{a^*}}. \end{aligned}$$

Kontrafaktuaaliset hasardit johdetaan vasteen tiheysfunktion avulla. Luonnollisten vaikutusten lausekkeissa (4) odotusarvot jollakin ajanhetkellä t vastaavat nyt vasteen tiheysfunktiota ajanhetkellä t . Kontrafaktuaalinen tiheysfunktio saadaan summaamalla tiheysfunktio yli mediaattorin jakauman ja siitä voidaan suoraviivaisesti johtaa myös kontrafaktuaalinen välttöfunktio, jolloin kontrafaktuaalinen hasardi saadaan näiden suhteena. VanderWeele [21] johtaa hasardiin kaavat normaalijakautuneen mediaattorin tilanteessa. Seuraavassa ne johdetaan samaan tapaan tämän tutkielman tilanteessa, jossa mediaattori on diskreetti muuttuja.

Käytetään vasteen kovariaateille ehdollistetulle tiheysfunktiolle hetkellä t merkintää $f_{T_a M_{a^*}}(t|\mathbf{w})$, kun altiste on asetettu arvoon a ja mediaattori on asetettu arvoon, jonka se saisi, jos altistelle olisi asetettu arvo a^* . Oletetaan, että luvussa 3.3 esitetyt identifioituvuusehdot C ovat voimassa, ja olkoon \mathbf{W}_1 jokin ehdon C3 toteuttava muuttujajoukko ja \mathbf{W}_2 jokin ehdon C4 toteuttava

muuttujajoukko. Tällöin

$$\begin{aligned} f_{T_a M_{a^*}}(t|\mathbf{w}_1, \mathbf{w}_2) &= \sum_m f(t|a, m, \mathbf{w}_2) P_M(m|a^*, \mathbf{w}_1) \\ &= \sum_m \lambda(t|0, 0, 0) e^{\gamma_1 a + \gamma_2 m + \gamma_3 a m + \gamma_4' \mathbf{w}_2} \exp\{-\Lambda(t|0, 0, 0) e^{\gamma_1 a + \gamma_2 m + \gamma_3 a m + \gamma_4' \mathbf{w}_2}\} \\ &\quad \times P_M(m|a^*, \mathbf{w}_1), \end{aligned}$$

$$S_{T_a M_{a^*}}(t|\mathbf{w}_1, \mathbf{w}_2) = \sum_m \exp\{-\Lambda(t|0, 0, 0) e^{\gamma_1 a + \gamma_2 m + \gamma_3 a m + \gamma_4' \mathbf{w}_2}\} P_M(m|a^*, \mathbf{w}_1),$$

jossa $S_{T_a M_{a^*}}(t|\mathbf{w}_1, \mathbf{w}_2)$ on vastaava välttöfunktio. Vastaava hasardi on tällöin

$$\begin{aligned} \lambda_{T_a M_{a^*}}(t|\mathbf{w}_1, \mathbf{w}_2) &= \frac{f_{T_a M_{a^*}}(t|\mathbf{w}_1, \mathbf{w}_2)}{S_{T_a M_{a^*}}(t|\mathbf{w}_1, \mathbf{w}_2)} \\ &= \lambda(t|0, 0, 0) e^{\gamma_1 a + \gamma_4' \mathbf{w}_2} \times \\ &\quad \underbrace{\frac{\sum_m e^{(\gamma_2 + \gamma_3 a)m} \exp\{-\Lambda(t|0, 0, 0) e^{\gamma_1 a + \gamma_2 m + \gamma_3 a m + \gamma_4' \mathbf{w}_2}\} P_M(m|a^*, \mathbf{w}_1)}{\sum_m \exp\{-\Lambda(t|0, 0, 0) e^{\gamma_1 a + \gamma_2 m + \gamma_3 a m + \gamma_4' \mathbf{w}_2}\} P_M(m|a^*, \mathbf{w}_1)}}_{r(t; a, a^*, \mathbf{w}_1, \mathbf{w}_2)}. \end{aligned}$$

Jos mediaattori on kolmiluokkainen ja päätetapahtuma voidaan olettaa niin harvinaiseksi, että $\Lambda(t|0, 0, 0) \approx 0$, saadaan

$$r(t; a, a^*, \mathbf{w}_1, \mathbf{w}_2) \approx \pi(0|a^*, \mathbf{w}_1) + e^{(\gamma_2 + \gamma_3 a)} \pi(1|a^*, \mathbf{w}_1) + e^{2(\gamma_2 + \gamma_3 a)} \pi(2|a^*, \mathbf{w}_1),$$

jossa $\pi(m|a^*, \mathbf{w}_1)$ merkitsee todennäköisyyttä $P_M(m|a^*, \mathbf{w}_1)$. Luonnolliset vaikutukset ovat tällöin

$$TE^{(HR)} \approx e^{\gamma_1(a-a^*)} \frac{\pi(0|a, \mathbf{w}_1) + e^{(\gamma_2 + \gamma_3 a)} \pi(1|a, \mathbf{w}_1) + e^{2(\gamma_2 + \gamma_3 a)} \pi(2|a, \mathbf{w}_1)}{\pi(0|a^*, \mathbf{w}_1) + e^{(\gamma_2 + \gamma_3 a^*)} \pi(1|a^*, \mathbf{w}_1) + e^{2(\gamma_2 + \gamma_3 a^*)} \pi(2|a^*, \mathbf{w}_1)}, \quad (7)$$

$$NDE^{(HR)} \approx e^{\gamma_1(a-a^*)} \frac{\pi(0|a^{(DE)}, \mathbf{w}_1) + e^{(\gamma_2 + \gamma_3 a)} \pi(1|a^{(DE)}, \mathbf{w}_1) + e^{2(\gamma_2 + \gamma_3 a)} \pi(2|a^{(DE)}, \mathbf{w}_1)}{\pi(0|a^{(DE)}, \mathbf{w}_1) + e^{(\gamma_2 + \gamma_3 a^*)} \pi(1|a^{(DE)}, \mathbf{w}_1) + e^{2(\gamma_2 + \gamma_3 a^*)} \pi(2|a^{(DE)}, \mathbf{w}_1)}, \quad (8)$$

$$NIE^{(HR)} \approx \frac{\pi(0|a, \mathbf{w}_1) + e^{(\gamma_2 + \gamma_3 a^{(IE)})} \pi(1|a, \mathbf{w}_1) + e^{2(\gamma_2 + \gamma_3 a^{(IE)})} \pi(2|a, \mathbf{w}_1)}{\pi(0|a^*, \mathbf{w}_1) + e^{(\gamma_2 + \gamma_3 a^{(IE)})} \pi(1|a^*, \mathbf{w}_1) + e^{2(\gamma_2 + \gamma_3 a^{(IE)})} \pi(2|a^*, \mathbf{w}_1)}. \quad (9)$$

Kaavojen täysin parametriset muodot riippuvat lopulta mediaattorille oletetavan multinomimallin parametrisoinnista.

Harvinaisen päätetapahtuman oletus tehtiin sen takia, että tällöin kumulatiivisen perustason hasardin termit supistuvat lausekkeista pois. Tämä on välttämätöntä erityisesti, jos estimointi perustuu Cox'in malliin, koska kumulatiivista perustason hasardia ei voida arvioida tuntematta perustason hasardifunktiota. Mikäli perustason hasardifunktio kuitenkin tunnettaisiin, lausekkeitä voitaisiin arvioida eksaktissa muodossaan, mutta koska kumulatiivinen hasardi on ajasta riippuva, myös luonnolliset vaikutukset riippuisivat tällöin ajasta.

4.3.2 Tulkinta

Logaritmisien hasardin skaalalla vasteen malli olisi additiivinen, jolloin tilanne olisi analoginen lineaaristen mallien yhteydessä esitettyyn tilanteeseen. Parametriset esitykset ovat logaritmisien hasardin tilanteessa huomattavasti monimutkaisemmat, kuin tavallisessa lineaaristen mallien tilanteessa, mutta se ei haittaa kausaalipäätelyä, koska luonnolliset vaikutukset on jo määritelty kontrafaktuaalisten vasteiden avulla ja niiden tulkinta ei perustu valittujen mallien parametreihin.

Luonnolliset vaikutukset voidaan johtaa logaritmiselle hasardille vastaavasti kuin ne johdettiin lineaaristen mallien tilanteessa. Edellä esitetyt hasardiskaalalla määritellyt vaikutukset vastaavat logaritmisien hasardin skaalan vaikutusten siirtämistä eksponenttiskaalalle, esimerkiksi

$$NDE^{(HR)} = \exp \left\{ \log (\lambda_{T_a M_a(DE)}) - \log (\lambda_{T_{a^*} M_a(DE)}) \right\} = \lambda_{T_a M_a(DE)} / \lambda_{T_{a^*} M_a(DE)}.$$

Samalla voidaan määritellä luonnollisten vaikutusten selittämät ja aiheuttamat vaikutukset valitsemalla niiden perustasot sopivasti.

5 Aineisto

Tutkimuksen aineisto koostuu terveydenhuollon rekistereistä kerätystä kohortista, jossa on 1410 NF1-potilasta ja 14017 verrokkia. Kohorttiin on valittu kaikki löydetyt ja varmistetut NF1-diagnoosit Suomen keskus- ja yliopistosairaaloiden (poislukien Ahvenanmaa) potilasrekisteritiedoista vuosilta 1987–2011. Jokaista kohortin tapausta kohden on valittu kymmenen verrokkia, jotka on kaltaistettu iän, sukupuolen ja asuinpaikkakunnan mukaan. Verrokkien

valinnassa on suljettu pois yksilöt, joiden ensimmäisen asteen sukulaisella on NF1-diagnosi. Kunkin NF1-potilaan ja potilaalle valittujen verrokkien kohorttiin liittymisen ajaksi on otettu ikä potilaan ensimmäisen NF1:een liittyvän sairaalakäynnin hetkellä. Yksilön kohortista poistumisaika on puolestaan ikä päätetapahtuman, kuoleman tai maastamuuton hetkellä, mikä tahansa on sattunut ensin. Yksilöiden seuranta on lopetettu viimeistään 1.1.2015.

Kiinnostuksen kohteena olevat päätetapahtumat, iskeemiset sydänsairausdiagnoosit, on kerätty Terveyden ja hyvinvoinnin laitoksen ylläpitämästä Hoitoilmoitusrekisteristä. Hoitoilmoituksissa on ilmoitettu hoitoon tulon ensisijainen diagnoosi ja korkeintaan 5 sivudiagnoosia. Diagnoosit on merkitty ICD-tautiluokitusjärjestelmän (*International Classification of Diseases*) mukaisesti. Vuosina 1987–1995 on käytetty ICD-luokituksen versiota 9 ja vuodesta 1996 eteenpäin versiota 10. Iskeemistä sydänsairausdiagnoosia merkitsevät diagnosikoodit ovat versiossa 10 I20–I25, ja versiossa 9 410–414-alkuiset, poislukien 414.1-alkuiset, koodit. Muut sensurointia merkitsevät tapahtumat eli kuolema tai maastamuutto on kerätty väestötietorekisteristä.

Sosioekonomisen statuksen mittarina on käytetty koulutusastetta, koska se on tavallisista mittareista paras ennustamaan elintapatekijöitä, jotka oletettavasti kuuluvat SES:n vaikutusmekanismeihin. Tiedot yksilöiden koulutusasteista ovat peräisin Tilastokeskukselta. Koulutusaste on mitattu ISCED-asteikolla (*International Standard Classification of Education*), ja SES on jaettu sen perusteella kolmeen luokkaan 0–2, 3–4 ja ≥ 5 , joista ensimmäinen vastaa perusastetta, toinen keskiastetta ja kolmas korkea-astetta. SES:n mittaushetkeksi on valittu 30 vuoden ikä, koska tämän jälkeen SES yleensä pysyy vakaana. Kohortista hylättiin yksilöt, joiden kohortista poistuminen tapahtui ennen 30 vuoden ikää, ja kohorttiin jääneiden yksilöiden seuranta-aika aloitettiin aikaisintaan 30 vuoden iästä. Jos yksilöltä puuttui tieto koulutusasteesta 30-vuotiaana, valittiin viimeisin tieto ennen 30 vuoden ikää tai aikaisin tieto 30 vuoden iän jälkeen.

Synnyinkunta on jaettu kahteen luokkaan Suomen ympäristökeskuksen maaseutu–kaupunki-luokituksen mukaisesti. Synnyinkuntamuuttuja on koodattu siten, että arvo 1 merkitsee synnyinkunnan luokittelua kaupungiksi ja 0 maaseuduksi. Ulkomailla syntyneiltä yksilöiltä syntymäkunta korvattiin aikaisimmalla löydetyllä havainnolla yksilön asuinkunnasta Suomessa.

Alkuperäisestä kohortista hylättiin lopulta 6155 yksilöä, jolloin lopulliseen kohorttiin jäi 817 NF1-potilasta ja 8455 verrokkia. Päätetapahtuman kohdanneita yksilöitä oli yhteensä 623, joista NF1-potilaita oli 62 ja verrokkeja 561.

Henkilöaikaa kertyi yhteensä 43 439 628 päivää, joista NF1-potilaat kerryttivät 3 481 116 päivää ja verrokkit 39 958 512 päivää. NF1-potilaista päätetapahtuman kohtasi siis 7.6 % ja päätetapahtumia 100 000 henkilöpäivää kohden havaittiin 1.8. Verrokeista päätetapahtuman kohtasi 6.6 % ja päätetapahtumia 100 000 henkilöpäivää kohden havaittiin 1.4.

6 Simulointitutkimus

Mediaatioanalyysin mukaisten vaikutusten estimoinnin ja estimaattien keskivirheiden laskemisen oikeellisuuden tarkistamiseksi simuloitiin 1000 aineistoa, joissa jokaisessa oli 10 000 yksilöä. Jokaista aineistoa kohden tehtiin 1000 bootstrap-otosta, joiden perusteella arvioitiin keskivirhettä. Bootstrap-keskivirheiden pitäisi tarkentua kohti 1000 simulointiaineiston pisteestimaattien otoshajontaa. Simulointiaineistoissa ovat muuttujina `age_stop`, `status`, `nf1`, `ses`, `aob` ja `pnf`, jotka tarkoittavat ikää seurannan lopetushetkellä, päätetapahtumaindikaattoria, NF1-statusta, sosioekonomista statusta, synnyinkunnan maaseutu-kaupunki-luokitusta ja ensimmäisen asteen sukulaisen NF1-statusta.

6.1 Muuttujat

Altiste ja taustamuuttujat. Todellisen aineiston mukaisesti NF1-tapauksiksi valittiin n. 10 % yksilöistä. Koska todellinen aineisto on kerätty siten, että verrokkien ensimmäisen asteen sukulaisilla ei ole NF1-diagnoosia, asetettiin verrokeille `pnf = 0`, ja NF1-tapauksille `pnf` arvottiin satunnaisesti. Ilmiön taustalle oletetussa kausaaligraafissa synnyinkunnalla ei ollut yhteyttä NF1-statukseen eikä ensimmäisen asteen sukulaisen NF1-statukseen, joten se valittiin kullekin yksilölle satunnaisesti. Muuttujat `aob` ja `pnf` täyttävät luvussa 3.3 esitellyt identifioituvuusoletukset C, ja estimointi toteutetaan näille muuttujille ehdollistettuna.

Mediaattori. Sosioekonominen status simuloitiin käyttäen R-ohjelmiston `mnet`-paketin `multinom`-funktion mukaista parametrisointia

$$\frac{P(\text{ses} = m | \text{nf1}, \text{aob}, \text{pnf})}{P(\text{ses} = 0 | \text{nf1}, \text{aob}, \text{pnf})} = \exp \{ \theta_0^{(m)} + \theta_1^{(m)} \times \text{nf1} + \theta_2^{(m)} \times \text{aob} + \theta_3^{(m)} \times \text{pnf} \},$$

jossa $m \in \{1, 2\}$.

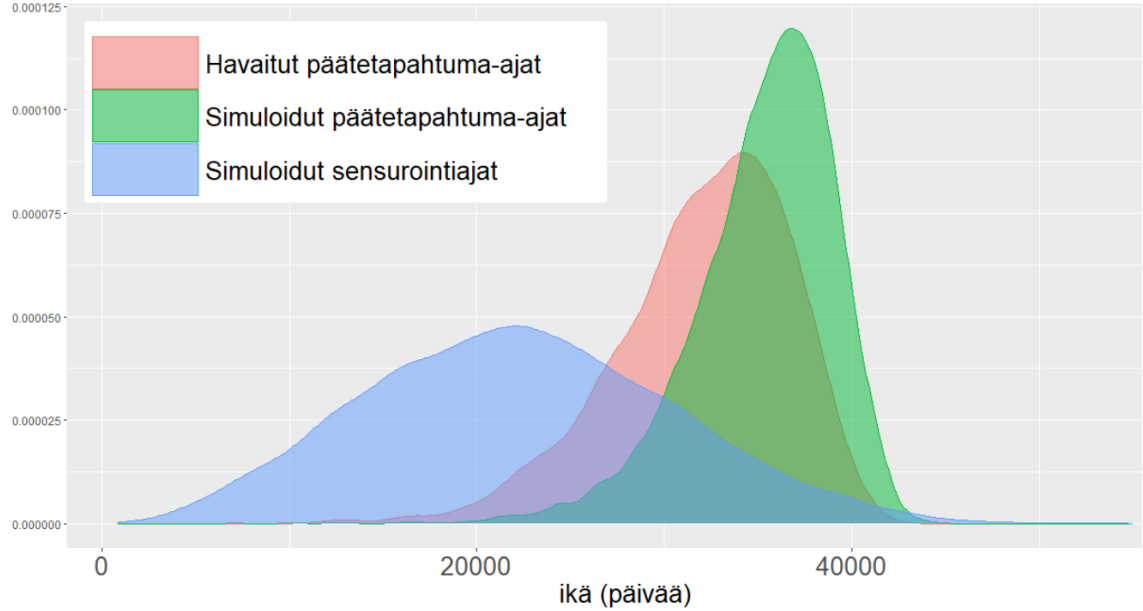
Mediaattorin malli			Vasteen malli	
	$m = 1$	$m = 2$		
$\exp\{\theta_0^{(m)}\}$	1.0	0.50	$\exp\{\gamma_1\}$	1.7
$\exp\{\theta_1^{(m)}\}$	0.90	0.10	$\exp\{\gamma_2\}$	0.70
$\exp\{\theta_2^{(m)}\}$	2.5	1.5	$\exp\{\gamma_3\}$	0.85
$\exp\{\theta_3^{(m)}\}$	0.89	0.33	$\exp\{\gamma_4\}$	0.90

Taulukko 1: Simulointikokeessa valitut mediaattorin ja vasteen mallien todelliset parametrien arvot. Mediaattorin mallia koskevat θ -parametreja vastaavat muuttujat ovat järjestyksessä: vakiotermi, NF1-status, synnyinkunta ja ensimmäisen asteen sukulaisen NF1-status. Kunkin parametrin arvo tarkoittaa kerrointa, jolla todennäköisyyksien $P(\text{SES} = m)$ ja $P(\text{SES} = 0)$ suhde muuttuu, kun parametria vastaavan muuttujan arvo muuttuisi nolasta yhteen. Vasteen mallin parametreja vastaavat muuttujat ovat järjestyksessä NF1-status, SES, synnyinkunta ja NF1–SES-yhdysvaikutus. Kunkin parametrin arvo tarkoittaa hasardisuhdetta kyseistä parametria vastaavan muuttujan luokkien välillä.

Mediaattorin mallin parametreille valitut todelliset arvot on esitetty taulukossa 1. Eksponentioidut parametrien arvot kuvaavat kerrointa, jolla parametria vastaavan muuttujan muutos yhden luokan ylöspäin vaikuttaa todennäköisyyksien $P(\text{ses} = m)$ ja $P(\text{ses} = 0)$ suhteeseen. Arvoltaan yli 1 olevat eksponentioidut parametrien arvot tarkoittavat, että kyseinen muuttuja kasvattaa todennäköisyyttä havaita $\text{ses} = m$ verrattuna $\text{ses} = 0$ ja alle 1 olevat arvot päinvastoin. Valitut arvot edustavat arvauksia siitä, millaisia tuloksia tämän tutkielman sovelluksessa saatettaisiin saada.

Vaste. Yksinkertaisuuden vuoksi simuloitiin ainoastaan iät seurannan päättymisen hetkellä ja oletettiin, että jokaisen yksilön seuranta olisi alkanut syntymästä. Yksilöille simuloitiin sensuroinnin ajat Weibull jakaumasta ja päätetapahtumien ajat Benderin ym. [23] mukaisesti Gompertz-jakaumaa approksimoivasta ääriarvojakaumasta, ja kohortista poistumisiäksi valittiin näistä varhaisempi. Jos yksilölle simuloitu päätetapahtuman aika oli aikaisemmin kuin sensuroinnin aika, yksilölle asetettiin status = 1 ja muuten status = 0. Kuvassa 5 on esitetty yhden simulointiaineiston jakaumat simuloituille sensurointihetkille ja päätetapahtumille sekä näiden perusteella saaduille havaituille

päätetapahtumille. Sensurointien ajat arvottiin jokaiselle yksilölle samasta jakaumasta, mutta päätetapahtuma-ajat riippuivat suhteellisten hasardien mallin muuttujille $nf1$, ses ja aob annetuista parametrien arvoista.



Kuva 5: Sensurointi- ja päätetapahtumaikien jakaumat kaikilta simuloituilta yksilöiltä yhdessä simulointiaineistossa. Simuloidut sensurointiajat ovat Weibull-jakaumasta ja simuloidut päätetapahtuma-ajat Gompertz-jakaumaa approksimoivasta ääriarvojakaumasta.

Simulointiaineistoissa päätetapahtumia oli keskimäärin 7.9 % kaikista yksilöistä, 9.9 % NF1-potilaista ja 7.6 % verrokeista. Päätetapahtumia miljoonaa henkilöpäivää kohden oli keskimäärin NF1-potilailla 4.5 ja verrokeilla 3.5. Simulointiaineistoissa saadut päätetapahtumat henkilöaikaa kohden ovat siis lähes kertaluokkaa pienemmät kuin todellisessa aineistossa, mikä johtuu siitä, että simulointiaineistoissa seuranta aloitettiin aina syntymähetkestä.

Suhteellisten hasardien malli parametrisoitiin seuraavasti:

$$\lambda(t|nf1, ses, aob) = \lambda(t|0, 0, 0) \times \exp \{ \gamma_1 \times nf1 + \gamma_2 \times ses + \gamma_3 \times aob + \gamma_4 \times nf1 \times ses \}.$$

Vastaaviksi todellisiksi parametreiksi valitut arvot on esitetty taulukossa 1. Kunkin parametrin γ_1 – γ_3 eksponentioidut arvot merkitsevät hasardisuhdetta vastaavan muuttujan luokkien välillä. Koska sosioekonominen status on kolmiluokkainen, parametrin arvo on hasardisuhde kahde peräkkäisen luokan välillä, eli tässä sosioekonomisen statuksen oletetaan olevan ns. välimatka-asteikollinen muuttuja. Yhdysvaikutuksen parametrin γ_4 eksponentti merkit-

see nfl-hazardisuhteiden suhdetta ryhmien $ses = j$ ja $ses = j - 1$ välillä. Arvoltaan yli 1 olevat eksponentioidut parametrin arvot tarkoittavat, että kyseisen muuttujan arvon kasvaminen yhdellä kasvattaa hasardia ja vastaavasti alle 1 olevat arvot tarkoittavat pienempää hasardia.

aob	pnf		TE	$NDE(0)$	$NDE(1)$	$NIE(0)$	$NIE(1)$
0	0	todellinen	0.563	0.469	0.487	0.0761	0.0939
		bs.est	0.565	0.475	0.489	0.0755	0.0903
		(tod.sd)	(0.236)	(0.214)	(0.239)	(0.0136)	(0.0307)
		(bs.sd)	(0.243)	(0.219)	(0.246)	(0.0139)	(0.0326)
1	0	todellinen	0.555	0.436	0.460	0.0954	0.119
		bs.est	0.559	0.443	0.464	0.0951	0.116
		(tod.sd)	(0.197)	(0.169)	(0.200)	(0.0157)	(0.0403)
		(bs.sd)	(0.203)	(0.172)	(0.206)	(0.0161)	(0.0422)
0	1	todellinen	0.551	0.477	0.491	0.0600	0.0738
		bs.est	0.554	0.481	0.493	0.0609	0.0726
		(tod.sd)	(0.244)	(0.225)	(0.245)	(0.0187)	(0.0312)
		(bs.sd)	(0.251)	(0.231)	(0.252)	(0.0192)	(0.0338)
1	1	todellinen	0.544	0.445	0.465	0.0790	0.0986
		bs.est	0.548	0.451	0.469	0.0798	0.0971
		(tod.sd)	(0.206)	(0.181)	(0.208)	(0.0224)	(0.041)
		(bs.sd)	(0.211)	(0.185)	(0.213)	(0.0227)	(0.0434)

Taulukko 2: Simulointikokeen luonnollisten vaikutusten logaritmiset estimaatit. Aineistoja simuloitiin 1000 ja jokaisessa oli 10 000 yksilöä. Riveillä ”bs.est” on ilmoitettu simuloitujen aineistojen perusteella laskettujen suorien (NDE) ja epäsuorien (NIE) vaikutusten piste-estimaattien keskiarvo ja riveillä ”todellinen” aineiston simuloinnissa asetettujen parametrien perusteella lasketut todelliset vaikutukset. Luonnolliset vaikutukset on laskettu kaikilla perustasojen valinnoilla ja käytetty perustaso on ilmoitettu vaikutuksen perässä sulkeissa. Lisäksi riveillä ”(tod.sd)” ja ”(bs.sd)” on ilmoitettu simuloitujen aineistojen piste-estimaattien otoshajonta ja toistoaineistojen bootstrap-keskivirheiden keskiarvot.

6.2 Simuloinnin tulokset

Olettaen, että päätetapahtuma on riittävän harvinainen, simulointikokeen todelliset luonnolliset vaikutukset saadaan sijoittamalla käytetyt todelliset parametrit luvussa 4.3.1 johdettuihin luonnollisten vaikutusten kaavoihin. Tuhannesta simulointiaineistosta laskettujen piste-estimaattien keskiarvojen pitäisi lähestyä todellisia vaikutuksia, ja bootstrap-keskivirheiden pitäisi lähestyä piste-estimaattien otoshajontaa.

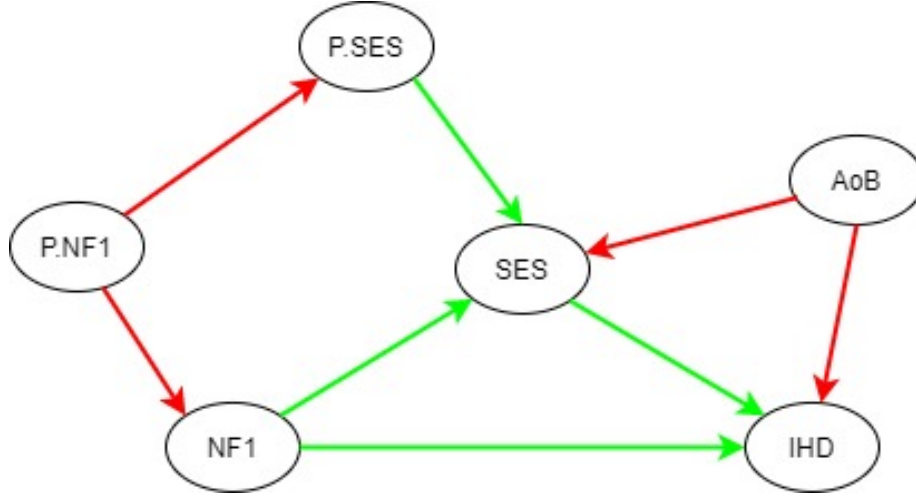
Estimaattien tarkkuuden arviointi keskivirheiden ja niistä laskettujen symmetristen luottamusvälien perusteella edellyttää estimaattorilta additiivista rakennetta. Toisin sanoen, suora keskivirheiden laskeminen ja niiden ilmoittaminen hasardisuhteen estimaatille ei ole mielekäästä, joten simulointikokeen yhteydessä estimoitavia vaikutuksia ja niiden keskivirheitä tutkitaan logaritmisin hasardin skaalalla.

Simulointiaineistojen piste-estimoinnin tulokset ja keskivirheet on ilmoitettu taulukossa 2. Taulukosta nähdään, että piste-estimaatit ovat keskivirheisiin nähden hyvin lähellä todellisia arvoja ja bootstrap-keskivirheet ovat hyvin lähellä piste-estimaattien otoshajontoja. Keskivirheet ovat kuitenkin vaikutusten suuruusluokkaan nähden melko suuria, mikä viittaisi, että tämän simulointikokeen mukaisissa tilanteissa luonnolliset vaikutukset eivät välttämättä estimoidu kovinkaan tarkasti.

7 Aineiston analysointi

Tässä luvussa sovelletaan luvussa 4.3 esitettyä hasardiskaalan parametrissa esitysmuotoa luonnollisten vaikutusten ja niiden keskivirheiden laskemiseksi. Käytettävä mediaatiomalli on jo tullut määritellyksi kuvassa 1. Tarkasteltava altiste on yksilön NF1-diagnoosi, mediaattori on yksilön kolmiluokkainen sosioekonominen status ja vaste on aika seurannan aloituksesta iskeemisen sydänsairauden diagnoosiin. Kuvassa 1 esitetyn kausaalirakenteen tilanteessa taustamuuttujat 'ensimmäisen asteen sukulaisen NF1-status' ja 'yksilön synnyinkunta' puhdistavat altisteen, mediaattorin ja vasteen väliset suhteet sekoittajista, joten niiden oletetaan täyttävän identifioituvuusehdot C. Kausaalirakenteessa on myös otettu huomioon vanhempien sosioekonominen status, joka on tuntematon muuttuja, mutta kuvassa 1 esitetyn kausaalirakenteen mukaan ensimmäisen asteen sukulaisen NF1-diagnoosin suhteen ehdollistaminen riittää taakaovipolun $NF1 \rightarrow P.NF1 \rightarrow P.SES \rightarrow SES$ sulkemiseen. Ensimmäisen asteen

sukulaisen NF1-diagnoosin ja yksilön synnyinkunnan suhteen ehdollistamisen jälkeen saatava kausaalirakenne on esitetty kuvassa 6. Vanhempien sosioekonomisella statuksella on yhä avoin vaikutus yksilön sosioekonomiseen statukseen, mutta se ei ole minkään suhteen sekoittaja.



Kuva 6: Muuttujien P.NF1 (ensimmäisen asteen sukulaisen NF1-status) ja AoB (yksilön synnyinkunta) suhteen ehdollistamisen vaikutus kausaalirakenteeseen. Punainen nuoli merkitsee, että polun kautta kulkeva vaikutus pidetään vakiona, ja vihreä nuoli merkitsee, että sen kautta kulkeva vaikutus on avoin muutoksille.

7.1 Mallioletukset

Suorien ja epäsuorien vaikutusten arviointi parametristen esitysten pohjalta edellyttää, että mediaattorille ja vasteelle määritellään mallit. Malleissa tarkasteltavat muuttujat on jo ennalta määrätty kausaaligraafissa, mutta vielä on tehtävä oletukset mallien parametrisista muodoista.

Mediaattori SES on diskreetti kolmiluokkainen muuttuja, joten oletetaan sille simulointikokeen (luku 6) yhteydessä esitelty multinomimalli:

$$\frac{P(\text{SES} = m | \text{NF1}, \text{AoB}, \text{P.NF1})}{P(\text{SES} = 0 | \text{NF1}, \text{AoB}, \text{P.NF1})} = \exp \{ \theta_0^{(m)} + \theta_1^{(m)} \times \text{NF1} + \theta_2^{(m)} \times \text{AoB} + \theta_3^{(m)} \times \text{P.NF1} \},$$

jossa $m \in \{1, 2\}$. Tällöin hasardiskaalan luonnollisten vaikutusten lausekkeissa tarvittavat mediaattorin ehdolliset todennäköisyydet ovat mallin parametrien

avulla esitettynä

$$P(\text{SES} = m | \text{NF1}, \text{AoB}, \text{P.NF1}) = \frac{\exp(\boldsymbol{\theta}^{(m)} \mathbf{X} \times I_0)}{1 + \exp(\boldsymbol{\theta}^{(1)} \mathbf{X}) + \exp(\boldsymbol{\theta}^{(2)} \mathbf{X})},$$

jossa

$$\boldsymbol{\theta}^{(m)} = \begin{pmatrix} \theta_0^{(m)} \\ \theta_1^{(m)} \\ \theta_2^{(m)} \\ \theta_3^{(m)} \end{pmatrix}^T, \quad \mathbf{X} = \begin{pmatrix} 1 \\ \text{NF1} \\ \text{AoB} \\ \text{P.NF1} \end{pmatrix} \quad \text{ja} \quad I_0 = \begin{cases} 0, & \text{jos } m = 0 \\ 1, & \text{jos } m \in \{1, 2\} \end{cases}.$$

Oletetaan vasteelle suhteellisten hasardien malli. Luvussa 4.3 johdettuisa luonnollisten vaikutusten lausekkeissa mediaattoria käsiteltiin diskreettinä välimatka-asteikollisena muuttujana. Tämä ei kuitenkaan välttämättä ole perusteltu oletus, joten parametrisoidaan malli siten, että

$$\lambda(t | \text{NF1}, \text{SES}, \text{AoB}) = \lambda(t | 0, 0, 0) \times \exp \{ \boldsymbol{\gamma} \mathbf{X} \},$$

jossa

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \gamma_6 \end{pmatrix}^T, \quad \mathbf{X} = \begin{pmatrix} \text{NF1} \\ I_1 \\ I_2 \\ \text{NF1} \times I_1 \\ \text{NF1} \times I_2 \\ \text{AoB} \end{pmatrix}, \quad I_1 = \begin{cases} 1, & \text{jos SES} = 1 \\ 0, & \text{muuten} \end{cases} \quad \text{ja} \quad I_2 = \begin{cases} 1, & \text{jos SES} = 2 \\ 0, & \text{muuten} \end{cases}.$$

Luonnollisten vaikutusten lausekkeissa (7)–(9) termit $\exp \{ (\gamma_2 + \gamma_3 \times \text{NF1}) m \}$ korvautuvat tällöin termeillä $\exp \{ (\gamma_2 + \gamma_4 \times \text{NF1}) I_1 + (\gamma_3 + \gamma_5 \times \text{NF1}) I_2 \}$.

7.2 Tulokset

Mediaattorille sovitettun multinomimallin ja vasteelle sovitettun Cox'in mallin tulokset on esitetty taulukossa 3. Mediaattorin mallin tuloksista nähdään, että NF1:llä vaikuttaisi olevan SES:iin negatiivinen vaikutus ja kaupungiksi luokiteltavalla synnyinkunnalla selvästi positiivinen vaikutus. Ensimmäisen asteen sukulaisen NF1-statuksen vaikutus puolestaan ei ole niin selkeä. Vasteen mallista nähdään, että piste-estimaattien valossa NF1 vaikuttaisi kasvattavan päätapahtuman eli iskeemisen sydänsairauden diagnoosin hasardia ja korkeampi SES puolestaan laskevan sitä. Yhdysvaikutustermeistä nähdään, että NF1:n

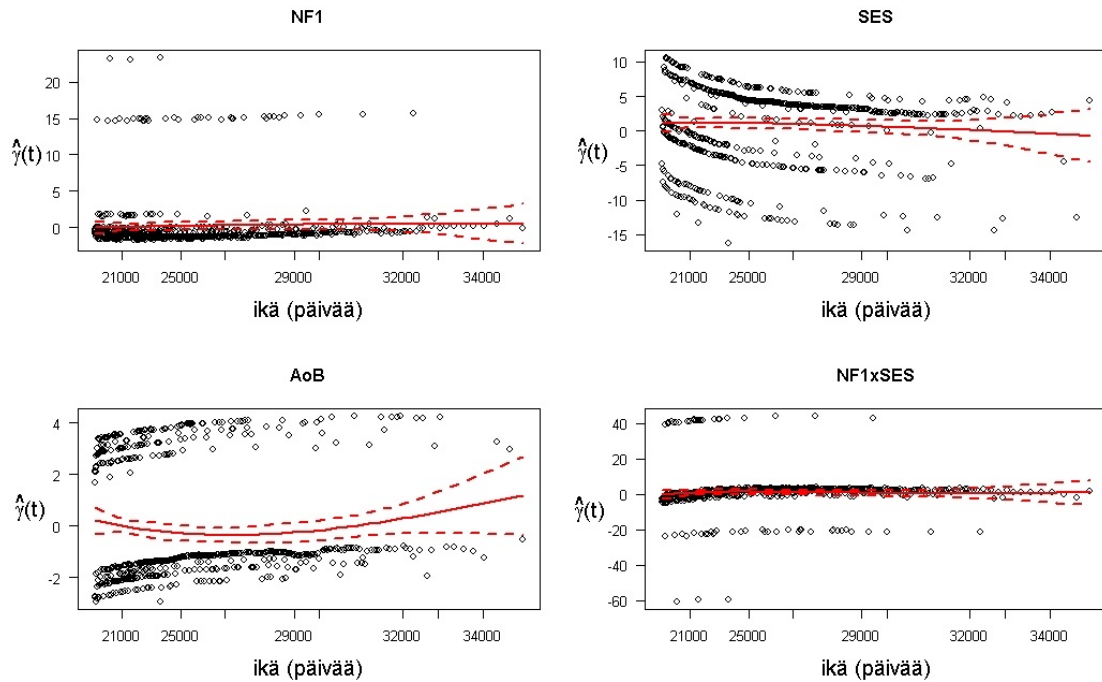
Mediaattorin malli			Vasteen malli			
	$m = 1$	$m = 2$				
$\exp\{\theta_0^{(m)}\}$	1.33 (1.24, 1.42)	0.87 (0.81, 0.94)	$\exp\{\gamma_1\}$	1.24 (0.87, 1.76)	$\exp\{\gamma_5\}$	1.79 (1.03, 3.13)
$\exp\{\theta_1^{(m)}\}$	0.87 (0.71, 1.06)	0.41 (0.32, 0.53)	$\exp\{\gamma_2\}$	0.75 (0.61, 0.93)	$\exp\{\gamma_6\}$	0.74 (0.22, 2.48)
$\exp\{\theta_2^{(m)}\}$	1.78 (1.60, 1.97)	2.39 (2.12, 2.69)	$\exp\{\gamma_3\}$	0.59 (0.46, 0.77)		
$\exp\{\theta_3^{(m)}\}$	1.37 (0.99, 1.90)	0.70 (0.43, 1.12)	$\exp\{\gamma_4\}$	0.89 (0.74, 1.06)		

Taulukko 3: Mediaattorille sovitetun multinomimallin ja vasteelle sovitetun Cox'in mallin estimaatit ja 95 %:n luottamusvälit. Mediaattorin mallissa parametri θ_0 on vakiotermi, θ_1 edustaa NF1-statuksen vaikutusta, θ_2 synnyinkunnan vaikutusta ja θ_3 ensimmäisen asteen sukulaisen NF1-statuksen vaikutusta. Vasteen mallissa parametreja vastaavat muuttujat ovat järjestyksessä: NF1-status, indikaattorimuuttuja SES = 1, indikaattorimuuttuja SES = 2, NF1-statuksen ja indikaattorimuuttujan 1 yhdysvaikutus, NF1-statuksen ja indikaattorimuuttujan 2 yhdysvaikutus ja synnyinkunta.

estimoitu vaikutus näyttäisi olevan hieman pienempi ositteessa SES = 1 kuin ositteessa SES = 0, mutta ositteessa SES = 2 NF1:n hasardia kasvattava vaikutus on puolestaan suurempi. Synnyinkunnan maasetutu-kaupunki-luokituksen vaikutuksen piste-estimaatti on alle 1, mutta luottamusväli hyvin suuri ja estimaatti siten epäluotettava.

Hasardien suhteellisuusoletuksen toteutumista arvioitiin Schoenfeldin residuaalien perusteella. Residuaalien lineaarisen ajasta riippuvuuden testin parvo koko mallille oli 0.28 ja muuttujille erikseen välillä 0.15–0.78. Schoenfeldin residuaalit kunkin päätetapahtuman hetkellä on esitetty kuvassa 7. Testin ja visuaalisten tarkastelujen perusteella hasardien suhteellisuusoletus vaikuttaa perustellulta.

Mediaattorin ja vasteen mallien parametriestimaattien perusteella lasketut luonnollisten vaikutusten estimaatit sekä bootstrap-keskivirheisiin perustuvat 95 %:n luottamusvälit on esitetty taulukossa 4. Taulukossa esitetyt estimaatit



Kuva 7: Schoenfeldin residuaalien tarkastelu Cox'in mallille. Punaiset viivat kuvaavat muuttujakohtaisia hasardisuhteen estimaatteja ajan funktiona ja punaiset katkoviivat niiden 95 %:n luottamusväliä. Hasardien suhteellisuusoletuksen pätiessä estimaattien tulisi olla vakioita, eli punaisen viivan tulisi olla likimain vaakasuora.

kuvaavat hasardisuhteita. Suoran vaikutuksen tulkitaan tarkoittavan kerrointa, jolla iskeemisen sydänsairauden hasardi sekoittavien tekijöiden ositteessa keskimäärin muuttuisi, jos verrokki muutettaisiin NF1-potilaaksi ilman, että se vaikuttaisi yksilön sosioekonomiseen statukseen. Epäsuora vaikutus puolestaan vastaa kerrointa, jolla verrokki–NF1-potilas-muutokseen liittyvä mediaattorin muutos keskimäärin vaikuttaisi hasardiin tarkasteltavassa ositteessa. Luonnolliset vaikutukset, joissa perustasoksi on valittu nolla, merkitsevät kyseisen luonnollisen vaikutuksen selittämää vaikutusta eli mittaavat sitä kausaalivaikutusta, joka voidaan selittää pelkällä kyseisen luonnollisen vaikutuksen edustamalla vaikutusmekanismilla. Vastaavasti luonnolliset vaikutukset, joissa perustasoksi on valittu yksi, merkitsevät kyseisen luonnollisen vaikutuksen aiheuttamaa vaikutusta eli mittaavat sitä kausaalivaikutusta, joka jäisi tapahtumatta ilman kyseisen luonnollisen vaikutuksen edustamaa vaikutusmekanismia. Totaalivaikutus koostuu nyt toisen vaikutuksen selittämästä ja toisen aiheuttamasta vaikutuksesta, mikä voidaan nähdä toteamalla, että kussakin

ositteessa $TE \approx NDE(0) \times NIE(1) \approx NDE(1) \times NIE(0)$. Täsmälliset yhtäsuuruudet eivät toteudu taulukon luvuilla pyöristysvirheiden takia.

AoB	P.NF1	TE	$NDE(0)$	$NDE(1)$	$NIE(0)$	$NIE(1)$
0	0	1.70 (1.27, 2.27)	1.56 (1.15, 2.13)	1.61 (1.21, 2.16)	1.05 (1.02, 1.09)	1.09 (1.00, 1.18)
1	0	1.78 (1.29, 2.46)	1.57 (1.10, 2.25)	1.68 (1.22, 2.32)	1.06 (1.02, 1.10)	1.13 (1.01, 1.27)
0	1	1.78 (1.31, 2.43)	1.69 (1.23, 2.33)	1.72 (1.26, 2.34)	1.04 (1.01, 1.06)	1.05 (0.99, 1.12)
1	1	1.88 (1.34, 2.64)	1.73 (1.22, 2.46)	1.80 (1.28, 2.53)	1.04 (1.01, 1.07)	1.09 (1.00, 1.18)

Taulukko 4: Luonnollisten vaikutusten piste-estimaatit ja 95 %:n luottamuskävyt sekoittavien tekijöiden eri ositteissa. Vaikutusten perässä sulkeissa on merkitty käytetty perustaso.

Taulukossa 4 esitetyt tulokset viittaavat siihen, että tyypin 1 neurofibromatoosilla on jokaisessa ositteessa selkeä kausaalinen vaikutus iskeemisen sydänsairauden hasardiin ja että kausaalivaikutus todella välittyy sekä suoran että epäsuoran vaikutuksen kautta. Lisäksi kummallakin luonnollisella vaikutuksella näyttäisi olevan hasardia kasvattava vaikutus. Epäsuoran vaikutuksen suuruus on kuitenkin hyvin pieni suoraan vaikutukseen verrattuna. Tämä on toisaalta odotettavaakin, koska epäsuora vaikutus mittaa tässä yhden spesifin vaikutusmekanismin kautta välittyvää kausaalivaikutusta, kun taas suora vaikutus sisältää kaikki muut vaikutusmekanismit. Tulosten perusteella voidaankin päätellä, että sosioekonomisen statuksen välittämä vaikutusmekanismi ei ole NF1-IHD-kausaalivaikutuksessa erityisen hallitseva mekanismi.

Perustasojen valinnalla ei näytä olevan suurta merkitystä luonnollisten vaikutusten estimaatteihin. Tämä viittaa siihen, että NF1-SES-yhdysvaikutus ei ole kovin merkittävä tekijä tarkasteltavassa ilmiössä. Estimoinnin tulokset eivät myöskään merkittävästi poikkea toisistaan sekoittavien tekijöiden eri ositteissa. Luonnollisia vaikutuksia koko populaation tasolla ei kuitenkaan voida päätellä pelkästään eri sekoittavien tekijöiden ositteiden perusteella.

8 Pohdinta

Tässä tutkielmassa esiteltiin modernin mediaatioanalyysin teoreettista perustaa ja näytettiin, miten luonnollisille kausaalivaikutuksille voidaan johtaa parametriset lausekkeet tapahtuma-aikavasteen ja diskreetin mediaattorin tilanteessa. Mediaatioanalyysin teoriaan liittyvä päätulos tutkielmassa oli luonnollisten vaikutusten parametrusten lausekkeiden johtaminen, kun vasteelle oletettiin suhteellisten hasardien malli ja mediaattorille multinomimalli. Näitä lausekkeitä ei tiettävästi ole aiemmin esitetty kirjallisuudessa. Lausekkeille esiteltiin parametrisoinnit sekä välimatka-asteikollisena että kategorisena muuttujana tarkasteltavan mediaattorin tilanteessa. Toisin kuin kirjallisuudessa yleensä, luonnollisten vaikutusten määrittelyissä perustasojen valinta jätettiin avoimeksi ja sen sijaan keskityttiin valinnan merkitykseen vaikutusten tulkinnaissa sekä esitettiin, miten eri valinnat suhteutuvat kirjallisuudessa usein käytettyihin määritelmiin. Lisäksi näytettiin, miten perustasojen valinnan merkitys ilmenee vaikutusten parametrusten esitysten yhteydessä.

Luonnollisten vaikutusten estimaattorien harhattomuus ja sitä kautta koko mediaatioanalyysin päättely perustuvat ilmiön taustalla olevan kausaalirakenteen olettamiseen. Mediaatiomallin arviointi kausaaligraafin perusteella on mielekästä, koska kausaaligraafien tulkinta on helppoa ja intuitiivista. Päättelyn perustaminen kausaaligraafiin on kuitenkin toisaalta ongelmallista, sillä mediaatioanalyysin prosessiin ei kuulu kausaalirakenteen oikeellisuuden testaaminen tai todistaminen. Imai ym. [24] suosittelee arvioimaan päättelyn herkkyyttä mahdollisille havaitsemattomille sekoittajille herkkyysanalyysin avulla ja esittää menetelmät herkkyysanalyysin toteuttamiseen useissa eri tilanteissa.

Luonnolliset vaikutukset halutaan usein estimoida parametrusten esitysten avulla. Parametrisen estimoinnin seurauksena kasvaa kuitenkin riski päätyä virheellisiin päätelmiin, sillä tällöin joudutaan tekemään lisäoletuksia mediaattorin ja vasteen parametruksista malleista. Luvussa 4.3 todettiin myös, että jos vasteelle oletetaan suhteellisten hasardien malli, joudutaan lisäksi tekemään oletus päätetapahtuman harvinaisuudesta. Suhteellisten hasardien mallin yhteydessä huomionarvoista on myös sen multiplikatiivisuus. Lange ja Hanssen [25] esittävät, että suhteellisten hasardien mallin matemaattinen rakenne soveltuu huonosti mediaatioanalyysiin, ja ehdottavat additiivisten hasardien mallia sen sijaan. Toisaalta VanderWeele näyttää, miten luonnolliset vaikutukset voidaan johtaa logaritmistien hasardien erotuksena [21], jolloin eksponen-

tioidut vaikutukset voidaan tulkita kausaalisisina vaikutuksina hasardisuhteen skaalalla.

Mediaatioanalyysin sovelluksena tutkittiin tyyppin 1 neurofibromatoosin kausaalivaikutusta iskeemisiin sydänsairauksiin. Luonnolliset vaikutukset estimoitiin käyttäen parametrisia esityksiä, jotka johdettiin mediaattorille ja vasteelle oletettujen parametrusten mallien pohjalta. Ilmiön taustalle oletettua kausaalirakennetta perusteltiin luvussa 2. Saadut tulokset viittaavat siihen, että NF1 kasvattaa iskeemisten sydänsairauksien hasardia ja että osa vaikutuksesta välittyy sosioekonomisen statuksen kautta. Lisäksi tulosten perusteella voitiin päätellä, että sosioekonomisen statuksen kautta välittyvä vaikutus ei ole koko NF1-IHD-vaikutussuhteen kannalta erityisen hallitseva vaikutusmekanismi. Cox'in mallin implikoima suhteellisten hasardien oletus vaikutti tarkasteluissa perustellulta oletukselta. Tulokset perustuivat lisäksi oletuksille päätetapahtuman harvinaisuudesta ja siitä, että kaikki tarpeelliset sekoittajat otettiin huomioon. Tulosten luotettavuutta voitaisiin parantaa tarkastelemalla tulosten herkkyyttä näiden oletusten rikkoutumiselle.

Viitteet

- [1] VanderWeele, T. (2016). *Mediation Analysis: A Practitioner's Guide*. Annual Review of Public Health, 37(1), 17–32.
- [2] MacKinnon, D. (2007). *Mediation analysis*. Annual Review of Psychology, 58(1), 593–614.
- [3] Hayes, A. (2009). *Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium*. Communication Monographs, 76(4), 408–420.
- [4] Alwin, D., & Hauser, R. (1975). *The Decomposition of Effects in Path Analysis*. American Sociological Review, 40(1), 37–47.
- [5] Fox, J. (1980). *Effect Analysis in Structural Equation Models: Extensions and Simplified Methods of Computation*. Sociological Methods & Research, 9(1), 3–28.
- [6] Judd, C., & Kenny, D. (1981). *Process Analysis: Estimating Mediation in Treatment Evaluations*. Evaluation Review, 5(5), 602–619.
- [7] Baron, R., & Kenny, D. (1986). *The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations*. Journal of Personality and Social Psychology, 51(6), 1173–1182.
- [8] Rubin, D. (1974). *Estimating causal effects of treatments in randomized and nonrandomized studies*. Journal of Educational Psychology, 66(5), 688–701.
- [9] Holland, P. (1988). *Causal Inference, Path Analysis, and Recursive Structural Equations Models*. Sociological Methodology, 18, 449–484.
- [10] Robins, J., & Greenland, S. (1992). *Identifiability and Exchangeability for Direct and Indirect Effects*. Epidemiology (Cambridge, Mass.), 3(2), 143–155.
- [11] Pearl, J. (2001). *Direct and indirect effects*. In Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence (UAI'01). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 411–420.

- [12] Imai, K., Keele, L., & Yamamoto, T. (2010). *Identification, Inference and Sensitivity Analysis for Causal Mediation Effects*. *Statistical Science*, 25(1), 51–71.
- [13] Pearl J. (2014). *Interpretation and identification of causal mediation*. *Psychological methods*, 19(4), 459–481.
- [14] Peltonen, S., Pöyhönen, M., Koillinen, H., Valanne, L., & Peltonen, J. (2014). *Miten tunnistan neurofibromatoosin?* *Duodecim*, 130(6), 619–625.
- [15] Kallionpää, R., Uusitalo, E., Peltonen, S., & Peltonen, J. (2017). *Neurofibromatoosi 1:n suuri syöpäalttius edellyttää valpasta seurantaa*. *Duodecim*, 133(20), 1877–1884.
- [16] Kallionpää, R., Peltonen, S., Leppävirta, J., Pöyhönen, M., Auranen, K., Järveläinen, H., & Peltonen, J. (2020). *Haploinsufficiency of the NF1 gene is associated with protection against diabetes*. *Journal of Medical Genetics*, Published Online First: 22 June 2020.
- [17] Kallionpää, R., Uusitalo, E., Leppävirta, J., Pöyhönen, M., Peltonen, S., & Peltonen, J. (2017). *Prevalence of neurofibromatosis type 1 in the Finnish population*. *Genetics in Medicine*, 20(9), 1082–1086.
- [18] Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd edition). Cambridge University Press.
- [19] VanderWeele, T. (2009). *Marginal Structural Models for the Estimation of Direct and Indirect Effects*. *Epidemiology (Cambridge, Mass.)*, 20(1), 18–26.
- [20] Pearl, J. (2012). *The Causal Mediation Formula—A Guide to the Assessment of Pathways and Mechanisms*. *Prevention Science*, 13(4), 426–436.
- [21] VanderWeele, T. (2011). *Causal Mediation Analysis With Survival Data*. *Epidemiology (Cambridge, Mass.)*, 22(4), 582–585.
- [22] Cox, D. (1972). *Regression Models and Life-Tables*. *Journal of the Royal Statistical Society. Series B, Methodological*, 34(2), 187–220.
- [23] Bender, R., Augustin, T., & Blettner, M. (2006). *Generating survival times to simulate Cox proportional hazards models by Ralf Bender, Thomas*

- Augustin and Maria Blettner*, *Statistics in Medicine* 2005; 24:1713–1723.
Statistics in Medicine, 25(11), 1978–1979.
- [24] Imai, K., Keele, L., & Tingley, D. (2010). *A general approach to causal mediation analysis*. *Psychological methods*, 15(4), 309–334.
- [25] Lange, T., & Hansen, J. V. (2011). *Direct and indirect effects in a survival context*. *Epidemiology (Cambridge, Mass.)*, 22(4), 575–581.