



**TURUN  
YLIOPISTO**

MENDELISTINEN SATUNNAISTAMINEN: KARDIOVASKULAARISTEN  
RISKITEKIJÖIDEN KAUSAALINEN VAIKUTUS KOGNITIIVISIIN  
TOIMINTOIHIN

Eetu Vento

Pro gradu -tutkielma  
Helmikuu 2025

Tarkastajat:  
Prof. Kari Auranen  
Dos. Juha Mykkänen  
Apulaisprof. Joni Virta

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

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

TURUN YLIOPISTO

Matematiikan ja tilastotieteen laitos

EETU VENTO: Mendelistinen satunnaistaminen: kardiovaskulaaristen riskitekijöiden kausaalinen vaikutus kognitiivisiin toimintoihin

Pro gradu -tutkielma, 35 s., 68 liites.

Tilastotiede

Helmikuu 2025

---

Tutkielmassa tarkastellaan kardiovaskulaaristen riskitekijöiden kausaalista vaikutusta kognitiivisiin toimintoihin. Kausaalivaikutuksen eli syy-seuraussuhteen estimoimiseksi käytettiin mendelististä satunnaistamismenetelmää, jossa havaintoaineistoon yhdistetään geneettistä tietoa muun muassa sekoittavien tekijöiden poissulkemiseksi. Geneettinen tieto toimii instrumenttimuuttujana, jonka avulla kausaalinen vaikutus voidaan estimoida. Tässä tutkielmassa geenitietona käytettiin polygeenistä riskisummaa, joka on muodostettu useasta altisteeseen eli riskitekijään vahvasti yhteydessä olevasta geenivariantista. Mendelistisessä satunnaistamisessa instrumenttimuuttujina käytettävien geenivarianttien idea perustuu siihen, että geenit ovat periytyneet vanhemmilta satunnaisesti, jolloin havainnoiva tutkimus voidaan muuttaa vastaamaan satunnaistettua kontrolloitua tutkimusta, kun instrumenttimuuttujilta vaadittavat ehdot täyttyvät.

Mendelistisinä satunnaistamismenetelminä käytettiin kaksivaiheista pienimmän neliösumman menetelmää sekä kolmiomenetelmää. Kaksivaiheisen pienimmän neliösumman menetelmän ensimmäisessä vaiheessa altisteelle ennustettiin polygeenisen riskisumman avulla arvot. Toisessa vaiheessa ennustetut altisteen arvot käytettiin selittävänä tekijänä vasteelle eli kognitiiviselle toiminnalle. Toisesta mallista saatava vaikutuksen estimaatti on altisteen kausaalinen vaikutus vasteeseen. Kolmiomenetelmässä polygeenisen riskisumman ja vasteen välinen yhteys jaettiin polygeenisen riskisumman ja altisteen välisellä yhteydellä, jolloin saadaan suoraan altisteen kausaalinen vaikutus vasteeseen.

Aineistona toimi Lasten Sepelvaltimotaudin Riskitekijät (LASERI) -tutkimuksen aineisto, jota aloitettiin kerätä jo vuonna 1980. Tällöin tutkimukseen valittiin satunnaisesti yhteensä 3596 tyttöä ja poikaa. Vuonna 2011 kognitiivista toimintaa mittaavat testit toteutettiin 2026 osallistujalle. Tutkielman vastemuuttujina käytettiin yhteensä viittä eri kognitiivista toimintaa mittaavaa testiä. Eri altistemuuttujia eli kardiovaskulaarisia riskitekijöitä valittiin tähän tutkimukseen myös viisi.

Tutkielman aihe pohjautuu aiemmin tehtyyn Rovion ym. tutkimukseen, jossa havaittiin usean lapsuudenaikaisen kardiovaskulaarisen riskitekijän yhteys keski-ikänsä kognitiiviseen toimintaan. Tässä tutkielmassa kaikkia näitä havaittuja yhteyksiä ei pystytty todentamaan kausaaliseksi. Useampaan aiemmin havaittuun yhteyteen voidaan siten olettaa vaikuttavan sekoittavia tekijöitä.

Asiasanat: mendelistinen satunnaistaminen, kausaalivaikutus, instrumenttimuuttuja



# Sisällys

<b>1 Johdanto</b>	<b>1</b>
<b>2 Instrumenttimuuttujista</b>	<b>3</b>
2.1 Kausaalisuus ja sekoittuminen . . . . .	3
2.1.1 Kontrafaktuaalit ja altisteen kausaalinen vaikutus . . . . .	3
2.1.2 Sekoittuminen . . . . .	4
2.2 Instrumenttimuuttajat ja kausaalinen päättely . . . . .	5
2.2.1 Binäärinen malli . . . . .	6
2.2.2 Esimerkki . . . . .	7
2.2.3 Kausaalivaikutuksen tulkinnasta . . . . .	9
2.2.4 Lineaarinen malli . . . . .	11
2.2.5 Mahdolliset ongelmat . . . . .	13
<b>3 Mendelistinen satunnaistaminen</b>	<b>14</b>
3.1 Geenivariantit ja niiden rooli mendelistisessä satunnaistamisessa . . . . .	14
3.2 Instrumenttimallien sovittaminen . . . . .	16
3.2.1 Kaksivaiheinen pienimmän neliösumman menetelmä . . . . .	16
3.2.2 Kolmiomenetelmä . . . . .	17
3.3 Kovariaattien käyttö . . . . .	18
3.4 Geneettisten instrumenttimuuttujien hyödyt . . . . .	19
<b>4 Aineisto ja tutkimuskysymykset</b>	<b>20</b>
4.1 Aineisto ja vastemuuttujat . . . . .	20
4.2 Altistemuuttujat . . . . .	21
4.3 Polygeeniset riskisummat . . . . .	21
4.4 Aineiston koko ja mallien implementointi . . . . .	22
<b>5 Tulokset</b>	<b>23</b>
5.1 Altisteen ja instrumenttimuuttujan välinen yhteys . . . . .	23
5.2 Altisteen ja vasteen välinen kausaalinen yhteys . . . . .	25
5.2.1 Verenpaine . . . . .	25
5.2.2 Kolesterolit . . . . .	28
5.2.3 Painoindeksi . . . . .	29
<b>6 Pohdinta</b>	<b>30</b>
6.1 Yhteenveto ja johtopäätökset . . . . .	31
6.2 Jatkotutkimuksen suuntia . . . . .	32
<b>Viitteet</b>	<b>34</b>
<b>A Liite: SAS-koodi</b>	<b>36</b>
<b>B Liite: R-koodi ja -tuloste</b>	<b>38</b>



# 1 Johdanto

Havainnoivat epidemiologiset tutkimukset ovat merkittävästi edistäneet ymmärtämme yleisistä sairauksista tunnistamalla eri sairauksien riskitekijöitä. Vaikka kausaalisuhteiden selvittäminen on keskeinen tavoite, niiden määrittäminen havainnoivien tutkimusten avulla on haastavaa. Oletettuja kausaalisuhteita onkin usein pyrittävä varmentamaan esimerkiksi satunnaistettujen kontrolloitujen kokeiden avulla. Erilaisten löydösten mahdollisia syitä voivat muun muassa olla havainnoivien tutkimusten havaitsemattomat sekoittavat tekijät, käänteinen kausaalisuus ja valikoitumisharha. Yksi keino sekoittavien tekijöiden sekä myös käänteisen kausaalisuuden poissulkemiseen on mendelistinen satunnaistaminen (*Mendelian randomization, MR*). Se on menetelmä, jota käytetään kausaalisuhteiden arvioimiseen yhdistämällä tutkimusaineistoon geneettistä tietoa tarkoituksena estimoida kausaalinen vaikutus eli syy-seuraussuhde havainnoivaa tutkimusta tarkemmin [1].

Mendelistinen satunnaistaminen käyttää hyödykseen instrumenttimuuttujina toimivia geneettisiä variantteja altisteen ja terveystavasteen kausaalisuhteen arvioimiseen. Geneettistä varianttia voidaan pitää instrumenttimuuttujana tietyille altisteille, jos se täyttää instrumenttimuuttujan oletukset: sillä on yhteys altisteeseen, se ei vaikuta vasteeseen muulla tavoin kuin altisteen kautta eikä se ole yhteydessä vasteeseen minkään sekoittavan tekijän vuoksi. [2]

Tämän tutkielman aihe pohjautuu Rovion ym. [3] tutkimukseen lapsuudenaikaisten kardiovaskulaaristen riskitekijöiden yhteydestä keski-ikäen kognitiivisiin toimintoihin. Aineisto perustuu laajaan LASERI-tutkimukseen (Lasten Sepelvaltimotaudin Riskitekijät), joka alkoi vuonna 1980 [4]. Kognitiiviset testit toteutettiin osalle tutkimusyksilöistä vuonna 2011. Tämän epidemiologisen havainnoivan seurantatutkimuksen tulokset viittasivat siihen, että useammalla kardiovaskulaarisella riskitekijällä olisi vaikutusta myöhempään kognitiiviseen toimintaan. Tutkimuksessa todettiin korkean verenpaineen, kohonneen kolesterolin sekä tupakanpolton vaikuttavan riippumattomasti ja heikentävästi keski-ikäen kognitiivisiin toimintoihin, erityisesti muistiin ja oppimiseen [3]. Kardiovaskulaarinen tarkoittaa sydämeen ja verisuoniin liittyvää. Kognitiivinen toiminta on tiedon käsittelyyn liittyvää toimintaa eli muun muassa havaitsemista, ajattelua ja muistamista.

Tämän työn tarkoituksena on omalla tavallaan toistaa Rovion ym. tutkimus käyttämällä apuna mendelististä satunnaistamismenetelmää. Tavoitteena on tutkia menetelmän käytettävyyttä sekoittavien tekijöiden kontrolloimiseksi eli tavoitteena on estimoida kardiovaskulaaristen riskitekijöiden kausaalista vaikutusta kognitiivisiin toimintoihin käyttämällä apuna instrumenttimuuttujia. Tutkielmassa altistemuuttujiin eli kardiovaskulaarisiin riskitekijöihin yhdistetään mendelistisen satunnaistamisen periaatteella geneettistä tietoa, jonka avulla pyritään sulkemaan sekoittavien tekijöiden vaikutusta pois eli arvioimaan altisteiden kausaalista vaikutusta vastemuuttujiin paremmin. Geneettisenä tietona käytetään snipeistä eli yksilön genomien tiettyjen yksittäisten kohtien varianteista laskettuja riskisummia, jotka toimivat eri altisteiden instrumenttimuuttujina. Snippi eli yhden nukleotidin polymorfismi/ yhden emäksen monimuotoisuus (*single nucleotide polymorphism, SNP*) on joitakin populaation yksilöitä koskeva muuntelu tietyssä DNA-sekvenssin kohdassa, jossa DNA-sekvenssin yksi emäs on korvautunut toisella emäksellä [5]. Mendelistisen

satunnaistamisen kehikossa geenivarianttien yhteydet altisteisiin ja muihin muuttujiin kuvastavat yhteyttä koko yksilön elinkaaren aikana, jolloin ei suoraan voida puhua enää lapsuudenaikaisten riskitekijöiden vaikutuksesta keski-ään kognitiivisiin toimintoihin, vaan puhutaan vain aiemmin havaittujen riskitekijöiden kausaalisesta vaikutuksesta kognitiiviseen toimintaan. Altistetta ja vastetta ei voida siis yhdistää suoraan tiettyyn ikävaiheeseen elinkaarella.

Tutkielman analyysivaiheessa hyödynnetään sekä kaksivaiheista pienimmän neliösumman menetelmää että suhde-estimaattoria, jota voidaan myös kutsua kolmiomenetelmäksi. Yksinkertaisesti esitettynä kaksivaiheisen pienimmän neliösumman menetelmän ensimmäisessä vaiheessa altistemuuttuja regressoidaan instrumenttimuuttujalla. Toisessa vaiheessa vastemuuttuja regressoidaan ensimmäisen vaiheen ennustetuilla altisteen arvoilla, jolloin saadaan estimaatti altisteen kausaaliselle vaikutukselle vasteeseen. Kolmiomenetelmässä sekä altiste- että vastemuuttuja regressoidaan instrumenttimuuttujan suhteen. Tämän jälkeen regressiomalleista saatavat kertoimet hyödynnetään siten, että instrumenttimuuttujan vaikutus vasteeseen jaetaan instrumenttimuuttujan vaikutuksella altisteeseen, jolloin ratkaisuna saadaan altistemuuttujan kausaalinen vaikutus vasteeseen.

Tutkielman luvussa 2 esitellään ja pohjustetaan instrumenttimuuttujien käytön ideaa sekä kausaalipäättelyn yleisiä lähtökohtia. Luvussa 3 perehdytään tarkemmin itse työn aiheeseen eli mendelistiseen satunnaistamiseen. Sitä seuraavassa luvussa 4 esitellään tutkielman aineisto sekä analyysivaiheessa käytettävät muuttujat. Oleellisimmat analyysien tulokset ovat luvussa 5. Luvussa 6 on yhteenveto, yleistä pohdintaa sekä ideoita jatkotutkimuksiin. Tutkielman lopun liitteistä löytyvät käytetyt ohjelmointikoodit sekä kaikki tulokset R Markdown -tulosteen muodossa.

## 2 Instrumenttimuuttujista

Tässä luvussa esitellään yleisesti instrumenttimuuttujamenetelmiä sekä niihin ja työhön oleellisesti liittyviä käsitteitä kuten kausaalisuutta ja sekoittumista. Instrumenttimuuttujamenetelmistä esitellään erilaisia perusmalleja sekä niihin liittyviä esimerkkejä ja tulkintoja.

### 2.1 Kausaalisuus ja sekoittuminen

Satunnaistetut ja kontrolloidut kokeet ovat lähtökohtaisesti paras tapa altisteen ja vasteen välisen vaikutuksen estimoimiseen lääketieteellisissä tutkimuksissa. Tämä ei kuitenkaan aina ole mahdollista, vaan usein on tukeuduttava havainnoivaan tutkimukseen. Esimerkiksi epidemiologiassa havainnoivat tutkimukset ovat yleisiä yrittäessä tutkia kausaalisia vaikutuksia eli syy-vaikutussuhteita. Vaikutuksen estimoinnin yleisin ongelma havainnoivissa tutkimuksissa on mahdollinen harha, joka syntyy sekoittumisesta eli joidenkin taustamuuttujien usein vahvastakin vaikutuksesta sekä altisteseen että vasteeseen. Näitä muuttujia kutsutaan sekoittaviksi tekijöiksi. Yleiset menetelmät kuten osittaminen tai jotkin regressiomenetelmät ovat usein hyviä vaihtoehtoja ratkaisemaan sekoittumisen ongelmaa silloin, kun sekoittavat tekijät ovat havaittuja. Usein sekoittavat tekijät ovat kuitenkin havaitsemattomia, jolloin potentiaalisena menetelmänä kausaalisten vaikutusten estimoimiseen voidaan pitää instrumenttimuuttujamenetelmää. [6]

#### 2.1.1 Kontrafaktuaalit ja altisteen kausaalinen vaikutus

Vaikka useimmissa tutkimuksissa yksilön altistus havaitaan ja sitä voidaan tutkia vain yhdellä tasolla, voidaan kausaalisten vaikutusten estimoinnissa tukeutua kontrafaktuaaliseen lähestymistapaan: mitä yksilön altisteen vaikutus vasteeseen olisi ollut altisteen eri tasoilla? Joissakin kokeellisissa asetelmissä kuten ristikkäistutkimuksissa, joissa yksilölle on mahdollista antaa kahta eri lääkettä, nämä eri altisteen tasot voidaan toteuttaa, mutta silloinkin eri aikapisteissä.

Määritettäessä yksilökohtainen kausaalinen vaikutus oletetaan, että halutaan arvioida yksilön altisteen vaikutusta vasteeseen kahdella eri altisteen tasolla. Kontrafaktuaalisessa päättelyssä tehdään oletus, että

1. Yksilön  $i$  altiste  $X$  olisi voinut olla kumpikin tasoista  $t$  tai  $c$  ( $X_{i,t}$  tai  $X_{i,c}$ ) ja
2. Vaste on mahdollinen kummallakin tasolla  $X_{i,t}$  ja  $X_{i,c}$ . [7]

Tosiasiassa vaste voidaan kuitenkin havaita vain yhdeltä altistuksen tasolta samanaikaisesti. Jos yksilön  $i$  altisteen taso on  $t$ ,  $Y_{i,c}$  on havaitsematon, ja sama toisinpäin. Sitä altisteen tasoa, jota yksilö ei saa, kutsutaan kontrafaktuaaliseksi altisteeksi. Tämän altisteen vastetta taas kutsutaan kontrafaktuaaliksi tai mahdolliseksi vasteeksi. Kontrafaktuaali pyrkii vastaamaan kysymykseen, mitä olisi tapahtunut, jos altiste olisi ollut eri tasolla. [7]

Jos altisteen tasolla  $t$  on eri vaikutus vasteeseen kuin altisteen tasolla  $c$  yksilöllä  $i$  eli  $Y_{i,t} \neq Y_{i,c}$ , altisteella on kausaalinen vaikutus vasteeseen. Tällöin yksilön altisteen eri tasojen vaikutusten erotus  $Y_{i,t} - Y_{i,c}$  on juurikin altisteen kausaalinen

vaikutus yksilössä  $i$ . Koska altisteen kausaalista vaikutusta vasteeseen ei voida tutkia samalla yksilöllä altisteen eri tasoilla, voidaan vain arvioida keskimääräistä kausaalista vaikutusta (*average causal effect*). Keskimääräisellä kausaalilla vaikutuksella tarkoitetaan yksilöiden keskimääräistä kausaalista vaikutusta kohdepopulaatiossa. [7]

Keskimääräinen kausaalinen vaikutus on odotusarvo

$$E[Y_{i,t} - Y_{i,c}].$$

Tämä voidaan myös esittää muodossa

$$E[Y_{i,t}] - E[Y_{i,c}],$$

jossa  $E[Y_{i,t}]$  on vasteen populaatiotason odotusarvo vasteen tasolla  $t$  ja  $E[Y_{i,c}]$  vasteen populaatiotason odotusarvo vasteen tasolla  $c$ . Tästä lähtien yksilöä merkitsevä indeksi  $i$  on jätetty merkitsemättä, vaikka estimoitavat kausaaliset vaikutukset ovatkin aina odotusarvoja yli yksilöllisten muuttujien ja muuttujat (esimerkiksi  $Y$ ) ovat yksilökohtaisia. Itse odotusarvot ovat populaatiotason arvoja, joiden erotuksiin kausaaliset vaikutukset perustuvat.

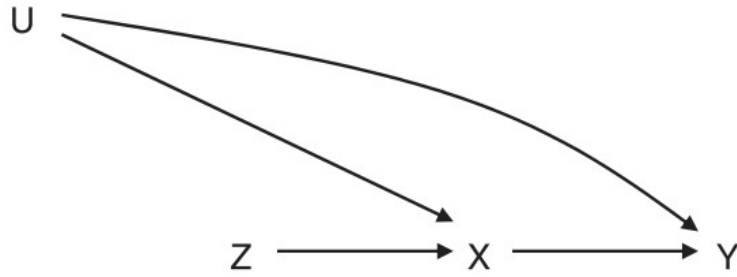
### 2.1.2 Sekoittuminen

Koska molempia altisteen arvoja ei voida havaita samalta yksilöltä, tulevat altisteen eri tasojen yksilöt käytännössä eri populaatioista. Oletetaan, että altisteen tason  $t$  yksilöt tulevat kohdepopulaatiosta  $A$  ja altisteen tason  $c$  yksilöt kontrollipopulaatiosta  $B$ . Oletetaan, että vasteen odotusarvo populaatiossa  $A$  on  $E[Y_{A,t}]$ , jos altistuksen taso on  $t$ , ja  $E[Y_{A,c}]$ , jos altistuksen taso on  $c$ . Altisteen eri tasoista johtuva kausaalinen vaikutus voidaan mitata odotusarvojen  $E[Y_{A,c}]$  ja  $E[Y_{A,t}]$  välisenä muutoksena eli erotuksena  $E[Y_{A,t}] - E[Y_{A,c}]$ . Kun kohdepopulaation  $A$  altistetaso on  $t$ , voidaan arvioida vain odotusarvoa  $E[Y_{A,t}]$ , jolloin  $E[Y_{A,c}]$  jää havaitsematta. Oletus onkin, että kontrollipopulaation  $B$  altistetaso  $c$  vasteen odotusarvo  $E[Y_{B,c}]$  on sama kuin  $E[Y_{A,c}]$  olisi ollut. Jos näin ei ole eli  $E[Y_{B,c}] \neq E[Y_{A,c}]$ , on kyseessä sekoittuminen. Tällöin populaatiot  $A$  ja  $B$  eroavat toisistaan altisteesta riippumattomista syistä, jotka ovat vastuussa odotusarvojen  $E[Y_{B,c}]$  ja  $E[Y_{A,c}]$  välisistä eroista. Populaatioiden  $A$  ja  $B$  vasteiden odotusarvojen erotus voidaan esittää muodossa

$$E[Y_{A,t}] - E[Y_{B,c}] = (E[Y_{A,t}] - E[Y_{A,c}]) + (E[Y_{A,c}] - E[Y_{B,c}]),$$

josta nähdään, että estimaatti  $E[Y_{A,t}] - E[Y_{B,c}]$  on sekoitus altisteen todellisesta eli kausaalisesta vaikutuksesta ( $E[Y_{A,t}] - E[Y_{A,c}]$ ) sekä virhetermistä eli harhasta ( $E[Y_{A,c}] - E[Y_{B,c}]$ ). Jos harhaa ei voida havaita, ei voida havaita myöskään todellista vaikutusta, mikä on usein tilanne epidemiologisissa tutkimuksissa. [8]

Yleisesti sekoittuminen määritellään siten, että on olemassa sekoittava(t) tekijä(t)  $U$ , joka vaikuttaa sekä altisteen että vasteen arvoon, jolloin altisteen todellista (kausaalista) vaikutusta vasteeseen ei voida estimoida tarkasti. Edellisessä kappaleessa mainitussa tapauksessa sekoittavana tekijänä  $U$  voidaan pitää populaatiota. Vaikka altisteen kausaalinen vaikutus eli  $E[Y_t] - E[Y_c]$  olisi sama molemmissa



Kuva 1: Instrumenttimuuttuja  $Z$ , altistemuuttuja  $X$ , vastemuuttuja  $Y$  ja sekoittavat tekijät  $U$ . Kuvan lähde: [9]

populaatioissa  $A$  ja  $B$ , estimoitaessa erotusta  $E[Y_{A,t}] - E[Y_{B,c}]$  saadaan kuitenkin sekoittunut estimaatti altisteen todelliselle vaikutukselle populaatioiden  $A$  ja  $B$  erojen vuoksi. Tämän takia tutkimuksissa populaation satunnaistaminen eri vertailuryhmiin on tärkeää. Sekoittava tekijä  $U$  voi kuitenkin olla mikä tahansa tekijä, joka vaikuttaa sekä altisteen että vasteen arvoihin ja siten sekoittaa niiden välistä kausaalista suhdetta, jota usein halutaan estimoida. Havaittavat ja sekoittavat tekijät  $U$  onkin aina pyrittävä sisällyttämään malleihin.

## 2.2 Instrumenttimuuttujat ja kausaalinen päättely

Instrumenttimuuttujan käyttö voi tulla kyseeseen silloin, kun tutkitaan altisteen kausaalista vaikutusta vasteeseen tilanteessa, jossa voi olla havaitsemattomia sekoittavia tekijöitä [9]. Instrumenttimuuttujia on käytetty jo 1920-luvulta asti, mutta pääosin ekonometriassa. Niitä on käytetty myös terveystieteissä, mutta melko vähän epidemiologiassa [10], jossa tutkitaan terveyteen liittyviä syitä, yhteyksiä ja esiintyvyyksiä väestötasolla. Instrumenttimuuttujien avulla on mahdollista toteuttaa mendelistinen satunnaistaminen, jota sovelletaan tässä työssä myöhemmin aineiston analysointiin.

Kuva 1 esittää tilannetta, jossa muuttujien  $X$  ja  $Y$  välistä suhdetta sekoittaa muuttuja  $U$ . Instrumenttimuuttujan  $Z$  avulla pyritään estimoimaan kiinnostuksen kohteena olevaa altistemuuttujan  $X$  kausaalista vaikutusta vastemuuttujaan  $Y$ . Instrumenttimuuttujan avulla pyritään minimoimaan altistemuuttujan ja vastemuuttujan välistä sekoittumista, jota ei välttämättä voida havaita, jotta altisteen kausaalinen vaikutus vasteeseen voidaan estimoida paremmin. Havaitsemattomia sekoittavia tekijöitä ei voida vakioda perinteisen regressiomallinnuksen tai osittamisen avulla, jolloin apuna käytetään instrumenttimuuttujaa. Instrumenttimuuttujamenetelmän käyttöön vaaditaan kuitenkin erilaisten ehtojen täyttymistä. [9]

Jotta muuttujaa  $Z$  voidaan käyttää instrumenttina, sen täytyy täyttää seuraavat kolme vaatimusta [10]:

1. Muuttujalla  $Z$  on vaikutus altistemuuttujaan  $X$ .
2. Muuttuja  $Z$  vaikuttaa vastemuuttujaan  $Y$  ainoastaan muuttujan  $X$  kautta. Tämä on sama asia kuin, että altistemuuttujan  $Y$  arvo ei riipu instrumenttimuuttujan  $Z$  arvosta, kun  $X$  ja  $U$  on kiinnitetty.

3. Muuttujalla  $Z$  ei ole yhteisiä sekoittavia tekijöitä vastemuuttujan  $Y$  eikä altistemuuttujan  $X$  kanssa eli se on riippumaton sekoittavista tekijöistä  $U$ .

Oletetaan alustavasti, että tässä työssä:

1. Ehto 1 täyttyy, koska instrumenttimuuttujana toimivat geenit vaikuttavat riskitekijöihin eli altistemuuttujaan.
2. Ehto 2 täyttyy, koska valitut geenit eivät suoraan vaikuta vasteeseen eli kognitiivisiin toimintoihin, vaan vaikutus tapahtuu pelkästään altisteen eli toteutuneiden riskitekijöiden kautta.
3. Ehto 3 täyttyy, koska geneilla ja vasteella ei ole yhteisiä sekoittavia tekijöitä, jotka vaikuttaisivat molempien arvoihin. Myöskään geneilla ja altistemuuttujalla ei ole yhteisiä sekoittavia tekijöitä.

Intuitiivisena lähtökohtana instrumenttimuuttujamenetelmille Greenland toteaa kiinnostuksen kohteena olevan  $X$ :n kausaalisen vaikutuksen  $Y$ :hyn olevan estimoitavissa instrumenttimuuttujan  $Z$  avulla [10]:

$$ASSO_{ZY} = ASSO_{ZX}ASSO_{XY},$$

jossa ASSO merkitsee alaindeksien mukaisten muuttujien välistä assosiaatiota eli yhteyttä. Ratkaisemalla yhtälöstä  $X$ :n ja  $Y$ :n välinen assosiaatio saadaan

$$ASSO_{XY} = \frac{ASSO_{ZY}}{ASSO_{ZX}}. \quad (1)$$

Lauseke (1) on ehtojen 1–3 vallitessa erityisen hyödyllinen silloin, kun juurikin  $X$ :n ja  $Y$ :n välinen suhde on sekoittunut havaitsemattomien kovariaattien vuoksi, mutta  $Z$ :n ja  $X$ :n sekä  $Z$ :n ja  $Y$ :n väliset suhteet eivät ole sekoittuneita [10]. Kovariaatilla tarkoitetaan muuttujaa, joka vaikuttaa vasteeseen, mutta joka ei ole yleisesti kiinnostuksen kohteena.

### 2.2.1 Binäärinen malli

Hernán esittelee instrumenttimuuttujan käyttöä tilanteessa, jossa sekä altiste  $X$  että instrumenttimuuttuja  $Z$  ovat dikotomisii eli niiden on mahdollista saada vain kahta eri arvoa [9]. Myös binäärisen mallin tapauksessa altisteen vaikutus vasteeseen voidaan estimoida käyttäen apuna instrumenttimuuttujaa  $Z$ , jonka avulla pystytään sulkemaan pois sekä muuttujiin  $X$  että  $Y$  vaikuttavia sekoittavia tekijöitä. Muuttujan  $X$  kausaalinen vaikutus muuttujaan  $Y$  voidaan nyt lausekkeen (1) mukaisesti estimoida lausekkeella

$$\frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[X|Z = 1] - E[X|Z = 0]}. \quad (2)$$

Yllä olevassa kaavassa (2) osoittajassa on instrumenttimuuttujan  $Z$  vaikutus  $Y$ :hyn eli vastemuuttujaan. Nimittäjässä on instrumenttimuuttujan  $Z$  vaikutus  $X$ :ään eli altistemuuttujaan. Jos  $X$  saisi aina saman arvon kuin  $Z$ , nimittäjän arvoksi tulisi 1, koska  $E[X|Z = 1] - E[X|Z = 0] = 1 - 0 = 1$ . Tällöin  $X$ :n yhteys  $Y$ :hyn saadaan suoraan estimoimalla osoittaja eli  $Z$ :n suhde  $Y$ :hyn, koska instrumenttimuuttujan arvo on täysin sama kuin altisteen arvo. Jos  $Z$ :n ja  $X$ :n välinen yhteys on heikkoa ja nimittäjä on lähellä nollaa, osoittajan merkitys kasvaa. Tällöin estimoitavasta suhteesta tulee myös todennäköisemmin harhaisempi estimaattorin multiplikaatiivisen luonteen vuoksi. Onkin tärkeää, että instrumenttimuuttuja  $Z$  ja altistemuuttuja  $X$  ovat mahdollisimman korreloituneita keskenään. [9]

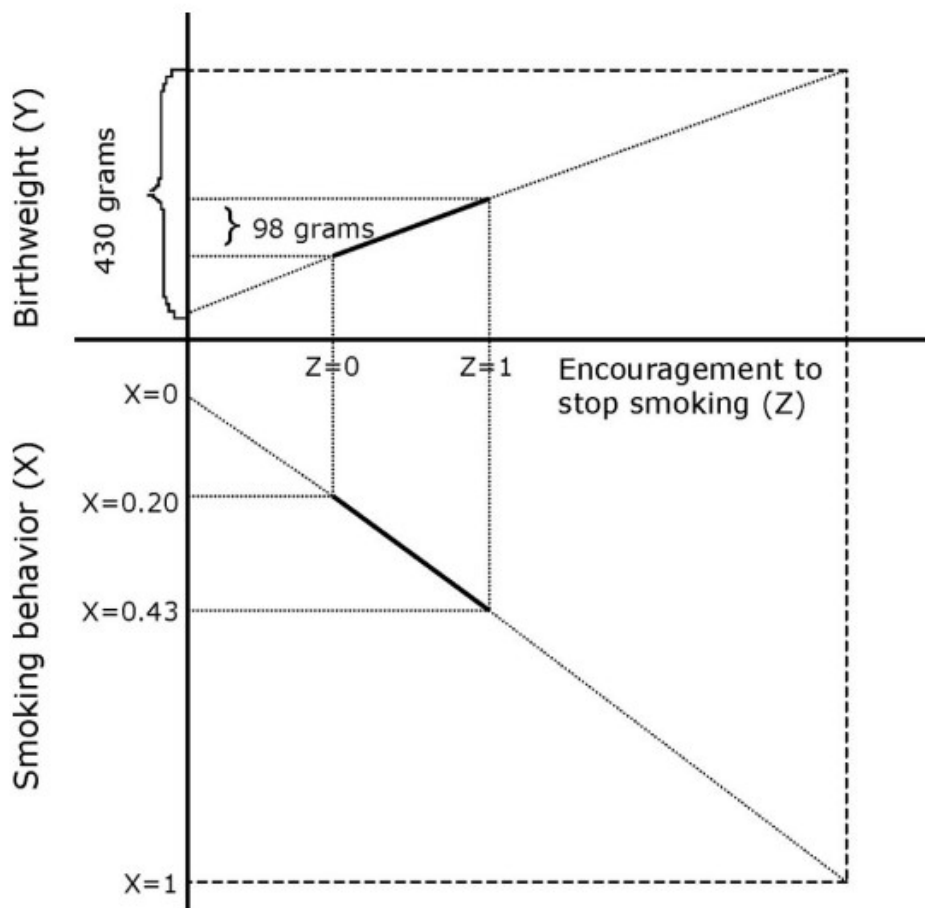
### 2.2.2 Esimerkki

Esitetään seuraavaksi esimerkki instrumenttimuuttujan käytöstä tapauksessa, jossa on binäärinen instrumentti- ja altistemuuttuja sekä jatkuva vastemuuttuja. Yksi varhaisimmista instrumenttimuuttujan käytöistä lääketieteellisellä alalla oli Permuttin ja Hebelin tutkimuksessa raskauden ajan tupakanpolton vaikutuksesta lapsen syntymäpainoon. Tutkimuksessa instrumenttimuuttujana toimi kehotus tupakanpolton lopettamiseen. [6]

Tutkimuksen tavoitteena oli tutkia altistemuuttujan ( $X$ ) eli tupakanpolton kausaalista vaikutusta vastemuuttujaan ( $Y$ ) eli lapsen syntymäpainoon. Instrumenttimuuttuja ( $Z$ ) oli kehotus tupakanpolton lopettamiseen, ja sen arvo määritettiin satunnaistamisen kautta. Satunnaistamisen ansiosta  $X$  ja  $Z$  olivat riippumattomia toisistaan. Instrumenttimuuttujalla  $Z$  eli kehotuksella tupakanpolton lopettamiseen voitiin myös olettaa olevan kausaalinen vaikutus altistemuuttujaan eli tupakanpolton lopettamiseen raskauden aikana. [6]

Instrumenttimuuttujan hyödyntämiseksi oli aluksi estimoitava  $Z$ :n ja  $Y$ :n eli instrumentin ja vasteen välinen yhteys. Tässä dikotomisen instrumenttimuuttujan tapauksessa estimaatti oli siten yksinkertaisesti lopettamaan kehotettujen ja ei-kehotettujen synnyttäneiden naisten lasten syntymäpainon keskiarvojen erotus. Tässä tutkimuksessa ero oli 98 grammaa enemmän painoa niillä lapsilla, joiden äitejä oli kehotettu lopettamaan tupakanpolto raskauden ajaksi. Seuraavaksi estimoitiin  $Z$ :n ja  $X$ :n välinen yhteys, jonka estimaatti saatiin tässä tapauksessa, kun tupakanpolton lopettamiseen kehotetuista laskettiin lopettaneiden osuus ja siitä vähennettiin kaikista ei-lopettamaan kehotetuista tupakanpolton lopettaneiden osuus. Lopettamaan kehotetuista 43 prosenttia lopetti oikeasti ja ei-kehotetuista 20 prosenttia, jolloin estimaatiksi saatiin  $0.43 - 0.20 = 0.23$ . Tämän erotuksen voidaan katsoa kattavan siten 23 prosenttia siitä vaikutuksesta, jossa kaikki lopettamaan kehotetut lopettaisivat ja kukaan ei-kehotetuista ei lopettaisi. Tästä johtuen oikean tupakanpolton vaikutuksen erotuksen syntymäpainoon voitiin estimoida olevan  $98g/0.23 \approx 426g$  polttavien ja ei-polttavien välillä. Idea tästä ajatuksesta ja estimaatin tulkinnasta on nähtävillä kuvasta 2. [6]

Martens muistuttaa, että vaikka instrumenttimuuttujan ja altisteen välinen riippuvuus on minimoitu kehotuksen satunnaistamisella, harhattomuuden oletus perustuu siihen ajatukseen, ettei kehotuksella tupakanpolton lopettamiseen ole vaikutusta muuhun kuin tupakanpoltoon, joka voisi vaikuttaa lapsen syntymäpainoon [6]. Kehotuksella voisi kuitenkin olla vaikutusta muihinkin terveyteen ja siten synty-



Kuva 2: Instrumenttimuuttujamenetelmän estimaatit Permuttin ja Hebelin tutkimuksessa. Lopettaneiden välinen 23 prosenttiyksikön ero lopettamaan kehotetuissa ja ei-kehotetuissa kattaa 23 prosenttia tupakanpolton lopettamisen kokonaisvaikutuksesta syntymäpainoon, jolloin ryhmien välinen estimoitu 98 gramman ero olisi todellisuudessa noin 430 grammaa lopettaneiden ja ei-lopettaneiden välillä. Kuvan lähde: [6]

mäpainoon vaikuttaviin asioihin, jolloin tupakanpolton vaikutusta syntymäpainoon oltaisiin yliestimoitu eli ero olisi oikeasti pienempi kuin estimoitu 426 grammaa pelkästään tupakanpoltosta johtuen. [6]

### 2.2.3 Kausaalivaikutuksen tulkinnasta

Kuten jo aiemmin todettiin, tavoitteena on yleisesti estimoida altisteen yksilökohtaista kausaalista vaikutusta vasteeseen altisteen eri tasoilla, mutta se ei ole saman yksilön kohdalla mahdollista, jolloin joudutaan tyytymään keskimääräisiin kausaalisiin vaikutuksiin. Tietyillä lisäoletuksilla, kolmen instrumenttimuuttujia koskevan ehdon lisäksi, voidaan kuitenkin osoittaa, että instrumenttimuuttujan avulla estimoitu kausaalinen vaikutus koskee koko populaatiota. Toisin sanoen tällöin kyseessä on vaikutus, joka vertaa tilannetta, jossa koko populaatio olisi altistettu, sellaiseen tilanteeseen, jossa koko populaatio olisi altistumaton. Ilman näitä lisäoletuksia kausaalinen vaikutus koskisi vain niitä, jotka tutkimuksessa olivat oikeasti altistuneet [9]. Tutkitaan tätä esimerkin kautta, jossa altisteet  $t$  ja  $c$  voivat olla vain kahdella eri tasolla, jolloin voidaan merkintöjen yksinkertaistamiseksi merkitä

$$Y_t = Y(1)$$

ja

$$Y_c = Y(0).$$

Tällöin altisteen yksilökohtainen kausaalinen vaikutus on

$$Y(1) - Y(0),$$

jossa sulussa oleva luku kertoo altisteen kontrafaktuaalisen tason. Koska molempia altisteen tasoja ei voida havaita samalta yksilöltä, estimoidaan koko populaation keskimääräistä kausaalista altisteen vaikutusta jo aiemmin esitettyyn tapaan:

$$E[Y(1)] - E[Y(0)].$$

Osoitetaan seuraavaksi, että yleinen instrumenttimuuttujien estimaattori estimoit altisteen vaikutusta altistuneilla kontrafaktuaalisessa mallissa, jossa  $Z$ :lla ei ole yhdysvaikutusta  $X$ :n kanssa. Tämä tarkoittaa sitä, että  $X$ :n vaikutus vasteeseen on sama riippumatta siitä, mikä  $Z$ :n arvo on. Myös  $Z$  on esimerkissä binäärinen. Esitetään myös muut tarvittavat lisäoletukset, jotta estimaattori sisältää myös keskimääräisen kausaalisen vaikutuksen  $E[Y(1)] - E[Y(0)]$  koko populaatiossa eli myös altistumattomissa. Seuraava esitys perustuu Hernánin artikkeliin [9].

Altisteen vaikutus altistuneissa ( $X = 1$ ) voidaan kirjoittaa

$$E[Y(1)|X = 1, Z] - E[Y(0)|X = 1, Z] = \psi_0 + \psi_1 Z,$$

tai yhtäpitävästi

$$E[Y|X, Z] - E[Y(0)|X, Z] = X(\psi_0 + \psi_1 Z), \quad (3)$$

jossa  $Y$  on vasteen havaittu arvo,  $\psi_0$  on altisteen vaikutus, jos  $Z = 0$ , ja  $\psi_0 + \psi_1$  altisteen vaikutus, jos  $Z = 1$ . Pelkkä  $\psi_0$  voidaan tulkita altisteen päävaikutuksena ja  $\psi_1$  altisteen ja instrumenttimuuttujan yhdysvaikutuksena vasteeseen.

Seuraavaksi oletetaan, että instrumenttimuuttujien ehdot 2 ja 3 täyttyvät eli instrumenttimuuttuja  $Z$  ja vastemuuttuja  $Y$  ovat riippumattomia toisistaan. Tällöin voidaan olettaa, että

$$E[Y(0)|Z = 1] = E[Y(0)|Z = 0] \quad (4)$$

eli muuttujan  $Y$  odotusarvo ei riipu muuttujan  $Z$  arvosta, jos altistetta ei ole/ei olisi. Ylläolevasta ehdosta (4) seuraa, että lauseke (3) voidaan esittää muodossa

$$E[Y - X(\psi_0 + \psi_1)|Z = 1] = E[Y - X\psi_0|Z = 0]. \quad (5)$$

Oletetaan myös, että  $\psi_1 = 0$  eli  $Z$ :n ja  $X$ :n välistä yhdysvaikutusta (ehto 2) vasteeseen  $Y$  ei ole, jolloin lauseke (5) voidaan kirjoittaa

$$E[Y|Z = 1] - E[X\psi_0|Z = 1] = E[Y|Z = 0] - E[X\psi_0|Z = 0]$$

$$\implies E[Y|Z = 1] - E[Y|Z = 0] = \psi_0(E[X|Z = 1] - E[X|Z = 0])$$

$$\implies \psi_0 = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[X|Z = 1] - E[X|Z = 0]}. \quad (6)$$

Tämä lauseke on täsmälleen samaa muotoa kuin aikaisemmin esitetty instrumenttimuuttujan binäärisen mallin lauseke (2). Esitetyillä oletuksilla osoitettiin siis se, että estimaattorin vaikutus  $\psi_0$  estimoii kausaalista vaikutusta vain niissä, jotka olivat altistuneet, ja vain, kun  $\psi_1 = 0$  eli kun yhdysvaikutusta instrumentin ja altisteen välillä ei ole. Jotta  $\psi_0$  estimoisi koko populaation keskimääräistä altisteen kausaalista vaikutusta myös altistumattomissa, on tehtävä vielä lisäoletuksia.

Oletetaan, että altisteen kausaalinen vaikutus on sama altistumattomilla ( $X = 0$ ) kuin altistuneilla ( $X = 1$ ) ja  $Z$ :lla ei myöskään ole yhdysvaikutusta vasteeseen altisteen  $X$  kanssa altisteen ollessa tasolla 0 eli

$$E[Y(1) - Y(0)|X = 0, Z = 1] = E[Y(1) - Y(0)|X = 0, Z = 0].$$

Tämä johtaa lisäoletukseen

$$E[Y(1) - Y(0)|X = 1] = E[Y(1) - Y(0)|X = 0].$$

Tällöin edellä johdettu parametrin  $\psi_0$  lauseke (6) estimoii altistuksen kausaalista vaikutusta koko populaatiossa. Tämä vaatii kummallakin altistuksen tasolla sen, ettei altisteella ole yhdysvaikutusta instrumenttimuuttujan kanssa eikä vasteen arvo riipu instrumenttimuuttujan arvosta. [9]

Jos malli oletetaan suoraan lineaariseksi samojen ehtojen täytyessä, voidaan se esittää muodossa

$$Y = \psi X + \epsilon.$$

Tällöin altistuneilla vaste olisi

$$Y(1) = \psi + \epsilon$$

ja altistumattomilla

$$Y(0) = \epsilon,$$

jolloin

$$E[Y(1)] - E[Y(0)] = \psi.$$

eli malli estimoi vaikutuksen yhtä lailla altistuneissa kuin altistumattomissakin. [9]

#### 2.2.4 Lineaarinen malli

Edellä instrumenttimuuttujien roolia esiteltiin lähinnä epäparametrisesti. Tarkastellaan seuraavaksi parametrissa lineaarista mallia ja muotoillaan instrumenttimuuttujan käyttö tässä yhteydessä. Oletetaan, että muuttujien  $Y$ ,  $X$  ja  $Z$  odotusarvojen riippuvuudet voidaan kuvata seuraavalla kolmella lineaarisella mallilla:

$$E(Y|X) = \gamma + \psi X, \tag{7}$$

$$E(X|Z) = \theta + \phi Z, \tag{8}$$

$$E(Y|Z) = \alpha + \beta Z. \tag{9}$$

Lausekkeiden (7) ja (8) perusteella muuttujan  $Y$  riippuvuus muuttujasta  $Z$  voidaan esittää seuraavasti:

$$\begin{aligned} E(Y|Z) &= E_{X|Z}(E_{Y|X,Z}[Y|X, Z]|Z) = E_{X|Z}(E_{Y|X}[Y|X]|Z) \\ &= E_{X|Z}[(\gamma + \psi X)|Z] = \gamma + \psi E[X|Z] \\ &= \gamma + \psi(\theta + \phi Z) \\ &= \gamma + \psi\theta + \psi\phi Z. \end{aligned}$$

Vertaamalla lausekkeeseen (9), saadaan

$$\begin{aligned} \alpha &= \gamma + \psi\theta, \\ \beta &= \psi\phi. \end{aligned}$$

Muuttujan  $X$  vaikutus muuttujaan  $Y$  saadaan ratkaisemalla

$$\psi = \frac{\beta}{\phi}, \tag{10}$$

jossa  $\beta$  kuvaa siis  $Z$ :n vaikutusta  $Y$ :hyn ja  $\phi$  kuvaa  $Z$ :n vaikutusta  $X$ :ään kuten lausekkeessa (1). Muuttujan  $X$  voidaan sanoa olevan endogeeninen, mikä tarkoittaa, että muuttuja voidaan selittää mallin muilla muuttujilla eli tässä tapauksessa

muuttujalla  $Z$ . Muuttuja  $Z$  taas on eksogeeninen eli sen arvo selittyy mallin ulkopuolisilla muuttujilla [6].

Martens esittää seuraavan kaavan asymptoottisesti harhattoman regressiokertoimen eli  $X$ :n vaikutuksen  $Y$ :hyn estimoimiseksi [6]:

$$\hat{\psi} = \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{n-1} \bigg/ \frac{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})}{n-1} = \frac{\hat{\sigma}_{Z,Y}}{\hat{\sigma}_{Z,X}},$$

jossa  $\hat{\sigma}_{Z,Y}$  on muuttujien  $Z$  ja  $Y$  välinen otoskovarianssi ja  $\hat{\sigma}_{Z,X}$  muuttujien  $Z$  ja  $X$  välinen otoskovarianssi. Tavanomaisempi tapa on esittää sama kahden regressiokertoimen pienimmän neliösumman lausekkeen eli PNS-estimaattorin suhteena:

$$\hat{\psi} = \frac{\hat{\sigma}_{Z,Y}}{\hat{\sigma}_{Z,X}} = \frac{\hat{\sigma}_{Z,Y}/\hat{\sigma}_Z^2}{\hat{\sigma}_{Z,X}/\hat{\sigma}_Z^2} = \frac{\hat{\psi}_{PNS(Z \rightarrow Y)}}{\hat{\psi}_{PNS(Z \rightarrow X)}},$$

jossa  $\hat{\sigma}_Z^2$  on instrumenttimuuttujan  $Z$  varianssi ja  $\hat{\psi}_{PNS}$  on pienimmän neliösumman estimaatti muuttujien välisiin vaikutuksiin liittyen. Osoittaja on siten instrumenttimuuttujan  $Z$  vaikutus vasteeseen  $Y$  ja nimittäjä instrumenttimuuttujan  $Z$  vaikutus altisteeseen  $X$ . Lausekkeen arvona saadaan siis ratkaistua altisteen  $X$  vaikutus vasteeseen  $Y$ , joka on mielenkiinnon kohteena. [6] Tämä voidaan esittää aikaisempaan tapaan (lauseke (10))

$$\hat{\psi} = \frac{\hat{\beta}}{\hat{\phi}}. \quad (11)$$

Martens esittää aikaisemmin esitetyt instrumenttimuuttujan ehdot seuraavasti lineaarisen mallin avulla hieman eri tavalla ilmaistuna [6]:

1. Muuttujien  $Z$  ja  $X$  on oltava korreloituneita, sillä muuten  $\hat{\psi}_{PNS(Z \rightarrow X)}$  on 0. Tällöin  $E(X|Z) = \theta$  eikä instrumenttimuuttujan käytöstä ole mitään hyötyä.
2. Instrumenttimuuttujalla  $Z$  ja lausekkeen  $Y = \gamma + \psi X + \varepsilon_Y$  virhetermillä  $\varepsilon_Y$  ei tulisi olla korrelaatiota. Tämä tarkoittaa sitä, ettei instrumenttimuuttuja saa vaikuttaa  $Y$ :n arvoon suoraan tai epäsuorasti muiden muuttujien (kuin  $X$ :n) kautta.
3. Instrumenttimuuttujalla  $Z$  ei saa olla sekoittavia tekijöitä altistemuuttujan  $X$  kanssa, jolloin lausekkeen  $X = \theta + \phi Z + \varepsilon_X$  virhetermin  $\varepsilon_X$  ja  $Z$ :n välillä ei tulisi olla korrelaatiota.

Nämä 3 ehtoa voidaan siten esittää muodossa [6]:

1.  $\rho_{Z,X} \neq 0$
2.  $\rho_{Z,\varepsilon_Y} = 0$
3.  $\rho_{Z,\varepsilon_X} = 0$ .

Sekoittavia tekijöitä  $U$  saa siis vain olla muuttujien  $X$  ja  $Y$  välillä. Nämä kolme ehtoa ovat oleellisesti samat ehdot kuin jo aikaisemminkin esitetyt.

### 2.2.5 Mahdolliset ongelmat

Instrumenttimuuttujien käyttöön liittyy mahdollisia ongelmia, jotka on syytä ottaa huomioon. Jos instrumenttimuuttujan ja altisteen välinen korrelaatio tai kausaalisuus on heikkoa, voi se johtaa suureen altisteen ja vasteen väliseen estimoituun harhaan, joka nähtiin jo lausekkeesta (2). Kuten yleisesti muissakin tilastollisissa menetelmissä, pieni otoskoko antaa myös instrumenttimuuttujamenetelmän tapauksessa harhaisempia tuloksia kuin tarpeeksi suurella otoskoolla tehdyt tutkimukset. Myös jos instrumenttimuuttujia koskevat ehdot eivät ole voimassa, voi se johtaa suureen harhaan. Erityisen huono tapaus on heikko instrumenttimuuttuja yhdistettynä pienen otoskokoon tai Martensin kolmannen ehdon rikkomiseen eli  $Z:n$  ja  $X:n$  välisten sekoittavien tekijöiden olemassaoloon. [6]

### 3 Mendelistinen satunnaistaminen

Mendelistinen satunnaistaminen on instrumenttimuuttujiin liittyvä erikoistapaus. Menetelmä perustuu olettamukseen, että jokin geneettinen variantti tai jotkin geneettiset variantit vaikuttavat vasteen arvoon vaikuttavaan altisteeseen, jonka taso saattaa muuten vaihdella ympäristöstä johtuen. Geneettisillä varianteilla on siten yhteys myös vasteen arvoon, jolloin variantteja voidaan käyttää instrumenttimuuttujina estimoitaessa altisteen kausaalista vaikutusta vasteeseen. Kuitenkin kuten instrumenttimuuttujan ehdoissa sanotaan, geneettisellä variantilla ei saa olla suoraa vaikutusta vasteeseen, vaan vaikutuksen on tapahduttava mielenkiinnon kohteena olevan altisteen kautta. [11]

Jos mendelistinen satunnaistamistutkimus tehdään pelkästään selvittämään, onko altisteella kausaalista vaikutusta vasteeseen, voi vaikutuksen suuruuden estimointi olla tarpeetonta. Jos taas halutaan tutkia vaikutuksen kvantitatiivista arvoa, tulevat myös muut asiat tärkeämmiksi. Näitä asioita voivat muun muassa olla, kuinka hyvin geneettinen variantti edustaa altistetta, ovatko geneettiset yhteydet altisteen kanssa estimoitu oikeanlaisesta populaatiosta ja ovatko muuttujien väliset suhteet lineaarisia ja homogeenisia populaatiossa. [2]

Mendelististen lakien mukaan instrumenttimuuttujina toimivien snippien eli yhden emäsparin vaihtelun geenivarianttien alleelit erottuvat toisistaan ja ovat satunnaisesti perittyjä vanhemmilta jälkikasvulle. Tämän periaatteen voidaan katsoa vastaavan satunnaistamista hoidolle satunnaistetussa kontrolloidussa kokeessa, mikä johtaa sekoittumattomaan altiste-vaste-suhteeseen. [1]

#### 3.1 Geenivariantit ja niiden rooli mendelistisessä satunnaistamisessa

Mendelistisessä satunnaistamistutkimuksessa altisteen kanssa vahvasti korreloivat snipit hyödynnetään instrumenttimuuttujina altisteen ja vasteen välisen kausaalisen vaikutuksen estimoimiseksi. Snipit noudattavat satunnaistamisen periaatetta, jolloin minimoidaan sekä sekoittavien tekijöiden vaikutus että mahdollinen käänteinen kausaalisuus, koska geneettiset variantit vaikuttavat sairauksien esiintymiseen, muttei toisinpäin. [12]

Viime vuosikymmeneltä lähtien on otettu suuria edistysaskelia geneettisten riskitekijöiden yhdistämisessä sairauksiin. Genominlaajuiset assosiaatiotutkimukset (*genome-wide association studies*, GWAS) ovat paljastaneet tuhansia geneettisiä, etupäässä snippeihin perustuvia yhteyksiä, joista yli 950 liittyy kardiovaskulaarisiin sairauksiin sekä mittauksiin (2018 huhtikuuhun mennessä). Näitä tietoja on myös julkisesti saatavilla [1]. Mendelistisessä satunnaistamistutkimuksessa yleinen tilastollinen lähestymistapa geenivarianttien valintaan on sisällyttää malliin kaikki variantit, jotka ovat tilastollisesti merkitsevästi yhteydessä altisteeseen. Usein käytetään genominlaajuisista merkitsevyyttä (*genome-wide significance*), jossa merkitsevyyden kynnyksarvona on  $p < 5 \cdot 10^{-8}$  [2]. Tilanteissa, joissa yksittäinen geenivariantti assosioituu voimakkaasti altisteeseen, voidaan tätä yksittäistä snippiä käyttää instrumenttina.

Heikkona instrumenttina pidetään instrumenttimuuttujaa, jonka tilastollisesti näytetty yhteys altistemuuttujaan ei ole vahva. Heikko instrumenttimuuttuja eroaa

kelvottomasta instrumenttimuuttujasta, joka ei täytä instrumenttimuuttujan ehtoja, siten, että heikon instrumenttimuuttujan avulla saatava estimaatti on asymp-toottisesti harhaton, jolloin heikkoa instrumenttimuuttujaa voidaan vahvistaa kasvattamalla otoskokoa. Instrumenttimuuttujan avulla saatavat estimaatit sisältävät systemaattista äärellistä otosharhaa, joka on yleensä havaitun (ja sekoittuneen) altisteen ja vasteen välisen yhteyden suuntaan. [5]

Altisteen ja instrumenttimuuttujan välisestä regressiomallista saatavaa F-testisuureen arvoa pidetään yleensä instrumenttimuuttujan vahvuuden mittana. Jos F-arvo on pienempi kuin 10, pidetään instrumenttimuuttujaa tavallisesti heikkona. Arvo on valittu, koska myöhemmin tarkemmin esiteltävän pienimmän neliösumman (*two-stage least squares*, 2SLS) menetelmän estimaatin harha, kun instrumenttimuuttujan F-testisuureen odotusarvo on 10, on rajoittunut 10 prosenttiin havaitun (ja sekoittuneen) vaikutuksen harhasta. Instrumenttimuuttujien binäärinen luokittelu heikkoihin ja vahvoihin on kuitenkin harhaanjohtavaa, koska harha on jatkuva ilmiö. Lisäksi odotettu harhan suuruus ei riipu kyseisen otoksen F-testisuureen arvosta vaan sen odotusarvosta. [5]

Mendelistisessä satunnaistamisessa geenivariantteihin pätevät aivan samat ehdot kuin muutenkin instrumenttimuuttujien kohdalla:

1. Geneettinen variantti on assosioitunut altisteen kanssa.
2. Geneettinen variantti on riippumaton (ehdolla altiste) vastemuuttujasta.
3. Geneettinen variantti on riippumaton sekoittavista tekijöistä.

Ensimmäinen ehto tarvitaan, koska mendelistisessä satunnaistamismenetelmässä (sekoittumaton) altiste estimoidaan käyttäen instrumenttimuuttujien alleelijakaunaa. Tämä ehto voidaan testata helposti ja todetaan hyväksytyksi, jos snippi-altiste-yhteyden F-testiarvo on yli 10. Toinen ehto on ekvivalentti sille, että instrumenttimuuttujalla ei ole vaikutusta vasteeseen, kun altisteen arvo pysyy kiinteänä. Tätä ehtoa on yleisesti hankala todistaa, koska snipeillä voi olla pleiotrooppisia vaikutuksia tai snipit ovat kytkentäepätasapainossa (*linkage equilibrium*) korreloituneita sellaisten geenien kanssa, joilla on vaikutusta vasteeseen altisteesta riippumattomasti. Pleiotropialla tarkoitetaan tilannetta, jossa geeni vaikuttaa useampaan yksilössä olevaan ominaisuuteen. Toisen ehdon rikkominen voi tapahtua myös, jos otos koostuu populaatiosta, jolla on eri alleelijakauma ja joka on myös assosioitunut vasteen kanssa. Tässä tapauksessa populaatorakenne on syynä sekä snipeille että vasteelle avaten polun snipistä vasteeseen, joka ei ole medioitunut altisteen kautta. Myös kolmatta ehtoa on vaikea varmentaa, koska samantapaisia ongelmia pleiotropian ja populaatorakenteen kanssa voi esiintyä liittyen altisteen ja vasteen välisiin sekoittaviin tekijöihin. [1]

Usean validin instrumenttimuuttujan käyttö auttaa käsittelemään heikon instrumentin aiheuttaman harhan ongelmaa. Lisäämällä useita ei-korreloituneita (kytkentätasapainoisia) snippejä esimerkiksi kaksivaiheisen pienimmän neliösumman malliin, voidaan lisätä tilastollista voimaa, mutta se saattaa myös lisätä suhteellista harhaa, jos lisätään heikkoja instrumentteja. Mendelistisen satunnaistamismenetelmän tilastollinen voima riippuu voimakkaasti siitä, minkä osuuden altisteen varianssista instrumenttimuuttuja selittää. Usean instrumenttimuuttujan käyttö, joko suorana

lisäyksenä tai alleelipistemääränä mallissa, voi siten lisätä tilastollista voimaa, koska suurempi osa altisteen varianssista on selitettynä. [1]

Jos kaikki geneettiset variantit ovat valideja ja kaikkien muuttujien (geneettiset variantit, altisteet ja vasteet) yhteydet ovat lineaarisia ja homogeenisia kaikilla yksilöillä populaatiossa, voidaan variantti-spesifien estimaattien odottaa kohdentuvan kaikki samaan kausaaliseen parametriin. Lisäksi variantti-spesifien estimaattien välillä ei oleteta olevan enempää heterogeenisuutta kuin mitä voitaisiin odottaa pelkästään sattumasta johtuvaksi. [2]

Tässä työssä instrumenttimuuttujana käytetään eräänlaista polygeenistä riskisummaa (*polygenic risk sum*, PRS), joka on muodostettu yhdistämällä usea altisteseen liittyvä snippi. Polygeenisen riskisumman käyttö on ekvivalenttia usean snipin painotetun keskiarvon käyttämiseksi instrumenttimuuttujana [13]. Polygeenisten riskisummien yhteydet muihin muuttujiin kuvastavat yhteyttä koko yksilön elinkaaren aikana. Polygeenisten riskisummien muodostamista on esitetty lyhyesti myöhemmin.

Mendelistiset satunnaistamistutkimukset voidaan toteuttaa käyttämällä yhden otoksen aineistoa (*one-sample Mendelian randomization*), jossa geneettiset variantit, altisteet ja vasteet on mitattu samoista yksilöistä, tai kahden otoksen aineistoa (*two-sample Mendelian randomization*), jossa varianttien ja altisteen yhteydet on estimoitu yhdestä aineistosta ja varianttien ja vasteen yhteydet toisesta aineistosta. Tässä työssä käytetään yhden otoksen aineistoa, jossa kunkin altiste-vaste-parin tutkimiseen käytetään yhtä instrumenttimuuttujaa eli kyseiseen altisteseen liittyvää polygeenistä riskisummaa.

## 3.2 Instrumenttimallien sovittaminen

Mendelistisen satunnaistamiseen liittyy useita tilastollisia menetelmiä, joiden käyttö vaihtelee hieman tilanteesta riippuen. Tässä tutkielmassa käytetään kaksivaiheista pienimmän neliösumman menetelmää sekä niin sanottua kolmiomenetelmää, jotka esitellään seuraavaksi. Nämä menetelmät ja niiden variaatiot ovat yleisimmät mendelistiset satunnaistamismenetelmät.

### 3.2.1 Kaksivaiheinen pienimmän neliösumman menetelmä

Yhden otoksen ja yksilötason aineiston tapauksessa voidaan kausaalinen vaikutus estimoida käyttäen kaksivaiheista pienimmän neliösumman menetelmää eli 2SLS-menetelmää. Ensimmäisessä vaiheessa altiste regressoidaan geneettisten varianttien ja mahdollisesti oleellisten kovariaattien suhteen. Toisessa vaiheessa vaste regressoidaan ensimmäisessä vaiheessa ennustettujen altisteen arvojen sekä samojen kovariaattien suhteen. Tarkkaan ottaen 2SLS-menetelmä tarkoittaa kaksivaiheista lineaarista regressioanalyysiä jatkuville altisteille ja vasteille [2]. Kaksivaiheinen pienimmän neliösumman menetelmä voidaan toteuttaa esimerkiksi tilasto-ohjelma R:n funktiolla `tsls`, joka löytyy paketista `sem` [1].

Koska tässä tutkielmassa otoksia on yksi, vastemuuttuja on jatkuva ja malli oletetaan lineaariseksi, voidaan aineiston analysoimiseksi käyttää kaksivaiheista pienimmän neliösumman menetelmää. Analyysit tehdään erikseen jokaiselle riskitekijälle, jolla on oma polygeeninen riskisummansa. Kaksivaiheisen pienimmän neliösumman menetelmän ensimmäisessä vaiheessa riskitekijä eli altistemuuttuja regressoi-

daan instrumenttimuuttujana toimivan polygeenisen riskisumman suhteen. Toisessa vaiheessa kognitiivisiin testeihin perustuvat vastemuuttujat regressoidaan erikseen ensimmäisestä regressiomallista saaduilla ennustetuilla altisteen arvoilla. Tästä toisesta mallista saatava altisteen regressiokerroin on tavoiteltu altisteen kausaalisen vaikutuksen estimaatti. [5]

Ensimmäisen vaiheen lineaarinen regressiomalli yksilölle  $i$  on muotoa

$$x_i = \alpha_0 + \alpha_1 z_i + \varepsilon_{X_i}.$$

Sovitetut eli ennustetut arvot  $\hat{x} = \hat{\alpha}_0 + \hat{\alpha}_1 z_i$  käytetään selittävänä muuttujana toisen vaiheen lineaarisessa regressiomallissa

$$y_i = \beta_0 + \beta_1 \hat{x}_i + \varepsilon_{Y_i}.$$

Malleissa  $\varepsilon_{X_i}$  ja  $\varepsilon_{Y_i}$  ovat riippumattomia virhetermejä ja mielenkiinnon kohteena oleva kausaalinen parametri on  $\beta_1$ . Jos molemmat mallit estimoidaan käyttäen tavanomaista pienimmän neliösumman regressiomenetelmää, molempien virhetermien oletetaan olevan homoskedastisia ja normaalijakautuneita. Homoskedastisuudella tarkoitetaan sitä, että toisella muuttujalla on suunnilleen sama varianssi kaikissa toisen muuttujan luokissa tai arvoissa. [5]

Vaikka 2SLS-menetelmän toinen vaihe antaa harhattoman piste-estimaatin kausaalille vaikutukselle, toisen vaiheen regressiomallista saatava keskivirhe ei ole oikein. Tämä johtuu siitä, ettei malli silloin ota huomioon ensimmäisen vaiheen regressiomallin epävarmuutta. Virhetermien homoskedastisuuden oletuksesta johtuen koko 2SLS-estimaattorin asympotoottinen varianssi on

$$\hat{\sigma}^2 (X^T Z (Z^T Z)^{-1} Z^T X)^{-1} = \hat{\sigma}^2 (\hat{X}^T \hat{X})^{-1},$$

jossa  $\hat{\sigma}^2$  on mallin  $y_i = \beta_0 + \beta_1 x_i + \varepsilon'_{Y_i}$  residuaalien varianssi,  $Z$  on instrumenttimuuttujien  $N \times 2$  matriisi (sisältää vakiotermin) ja  $X$  on altistemuuttujien  $N \times 2$  matriisi (sisältää vakiotermin). Menetelmästä saadun piste-estimaatin harha jatkuvan vasteen tapauksessa on approksimatiivisesti  $1/E(F)$  havaitun (sekoittuneen) yhteyden harhasta, jossa  $E(F)$  on F-testisuureen odotusarvo ensimmäisen vaiheen regressiomallista. [5]

### 3.2.2 Kolmiomenetelmä

Vaihtoehtoisesti kausaalinen vaikutus voidaan estimoida käyttämällä niin sanottua kolmiomenetelmää ilman tarvetta laskea ensin ennustettuja altisteen arvoja ja sen jälkeen niiden regressiokerrointa vasteeseen. Kolmiomenetelmän mukaan instrumenttimuuttujan suora vaikutus vasteeseen on kahden polulla olevan vaikutuksen tulo, jossa mediaattorina instrumenttimuuttujan ja vasteen välillä toimii altiste, jolloin siis kaavan (11) mukaan

$$\hat{\beta} = \hat{\psi} * \hat{\phi}.$$

Altisteen kausaalinen vaikutus vasteeseen on siten

$$\hat{\psi} = \frac{\hat{\beta}}{\hat{\phi}},$$

jossa  $\hat{\beta}$  on  $Z$ :n ja  $Y$ :n välinen suhde,  $\hat{\phi}$   $Z$ :n ja  $X$ :n välinen suhde ja näiden suhteena saatava  $\hat{\psi}$   $X$ :n ja  $Y$ :n välinen kausaalinen suhde. Kolmiomenetelmästä saatava kausaalinen vaikutus tunnetaan myös nimellä suhde-estimaatti (*ratio estimate*) tai Waldin estimaatti. Tätä menetelmää käytetään usein, jos käytetään kahta eri aineistoa. Tällöin kolmiomenetelmää voi hyödyntää R:ssä käyttämällä esimerkiksi R-pakettia TwoSampleMR. [1]

Kolmiomenetelmä, jota kutsutaan myös kertoimien suhde -menetelmäksi, on lähtökohtaisesti yksinkertaisin keino estimoida altisteen kausaalinen vaikutus vasteeseen. Menetelmän luottamusväli saadaan laskettua delta-menetelmällä [5]. Kolmiomenetelmästä saatavan regressiokertoimen  $\psi_{XY}$  keskivirheen SE estimoiva delta-menetelmä perustuu Taylorin sarjakehitelmään, ja se voidaan laskea approksimatiivisesti käyttäen kaavaa:

$$\text{var}(\psi) = \text{var}\left(\frac{\beta}{\phi}\right) \approx \frac{\text{var}(\beta)}{\phi^2} + \frac{\beta^2}{\phi^4}\text{var}(\phi) - 2\frac{\beta}{\phi^3}\text{cov}(\beta, \phi),$$

jossa  $\text{cov}(\beta, \phi)$  on kahden vaikutuksen kovarianssi. Tämä termi häviää, jos vaikutusten estimaatit on laskettu eri otoksista [1]. Kun keskivirhe  $SE$  on ensin laskettu, esimerkiksi kausaalisen vaikutuksen piste-estimaatin 95 prosentin luottamusväli saadaan estimoitua perinteiseen tapaan kaavalla  $\psi \pm 1.96 * SE$ .

Kolmiomenetelmän laajennus on painotettu käänteisen varianssin menetelmä (*inverse-variance weighted, IVW*), jossa käytetään samanaikaisesti useaa instrumenttimuuttujaa. Jokaisen instrumenttimuuttujana toimivan geneettisen variantin kausaalivaikutusestimaatit yhdistetään siten, että IVW-testi on geneettisten varianttien kausaalivaikutusten painotettu keskiarvo [14]. Yleisesti IVW-menetelmä antaa tarkentuvan arvion altistuksen kausaalivaikutuksesta vasteeseen, kun jokainen käytettävä geneettinen variantti täyttää instrumenttimuuttujan oletukset [15].

Jos vasteeseen vaikuttaisi usea altiste ja altisteilla on yhteisiä geneettisiä ennustajia, olisi turvaututtava toiseen lähestymistapaan: usean muuttujan mendelistiseen satunnaistamiseen (*multivariable Mendelian randomization*). Tämä menetelmä on yhden muuttujan mallin laajennus, jossa sallitaan samojen geneettisten varianttien olla assosioitunut useampaan kuin yhteen altistemuuttujaan. Usean muuttujan malli estimoii suoraa kausaalista vaikutusta jokaisen altistemuuttujan osalta yhdessä mallissa. [2]

### 3.3 Kovariaattien käyttö

Jos on mahdollista käyttää havaittuja kovariaatteja, jotka selittävät altisteen tai jatkuvan vasteen variaatiota, voidaan ne sisällyttää analyysiin. Kovariaattien sisällyttäminen lisää yleisesti tilastollista voimaa ja siten kausaalisen estimaatin tarkkuutta. Toisaalta se voi johtaa kausaalisen estimaatin harhaan, jos kovariaatti on altisteen ja vasteen välisellä kausaalisella polulla tai on törmäytinmuuttuja (*collider*) tai kausaalisesti törmäytinmuuttujan suuntaan. Törmäytin on muuttuja, johon sekä altiste että vaste vaikuttavat. Harhaa voi myös syntyä, jos kovariaatin sisältämä analyysimalli on määriteltä väärin. Kahden vaiheen estimoinnissa jokainen ensimmäisen regressiomallin kovariaatti täytyy sisällyttää myös toisen vaiheen regressiomalliin. Tämän tekemättä jättäminen voi aiheuttaa assosiaatiota instrumenttimuuttujien ja sekoittavien tekijöiden välille aiheuttaen harhaa. [5]

Joissakin tapauksissa kovariaattien suhteen vakiointi on tarpeellista varmistamaan instrumenttimuuttujien validius, jos instrumenttimuuttujien oletukset pitävät vain ehdollisesti kovariaattien suhteen. Esimerkki mendelistisen satunnaistamisen tapauksessa olisi populaation osittamisen tapaus, jossa otospopulaatio koostuu osapopulaatioista, joilla on eri jakaumat instrumenttimuuttujien ja vasteen suhteen. Nämä osapopulaatiot voivat vastata esimerkiksi eri etnisiä populaation osajoukkoja. Tällöin instrumenttimuuttujan ja vasteen välinen yhteys saattaisi ainoastaan vastata erilaisuuksia etnisyyksissä eikä biologisissa altisteen vaikutuksissa. Tämä voidaan käsitellä ainakin osittain genomilaaajuisella aineistolla vakioimalla geneettisten pääkomponenttien suhteen instrumenttimuuttuja-analyysissä, vaikka ei ole teoreettista varmuutta siitä, että geneettisten pääkomponenttien suhteen adjustoiminen ratkaisee populaation osittamisen ongelman. Yleisesti näiden tapausten ohella kehoitetaan adjustoimaan kovariaattien suhteen pääanalyysin tuloksissa arvioitaessa kausaalista vaikutusta [5]. Tässä tutkielmassa kovariaatteja ei ole lisätty malleihin.

### 3.4 Geneettisten instrumenttimuuttujien hyödyt

Mendelistiseen satunnaistamiseen liittyy useita hyötyjä verrattuna havainnoivaan tutkimukseen. Näitä ovat esimerkiksi [11]:

**Sekoittumisen minimointi.** Käyttämällä geneettistä varianttia ympäristön mahdollisesti muokkaaman altisteen edustajana, on epätodennäköistä, että sekoittumista olisi niin paljon kuin silloin, jos mitattaisiin altisteen vaikutusta pelkästään suoraan altisteesta vasteeseen.

**Käänteisen kausaalisuuden minimointi.** Päätelmät havainnoivista tutkimuksista voivat olla alttiita käänteisen kausaalisuuden harhalle. Tällöin vaste eli usein sairaus vaikuttaakin altistetasoihin. Jotkut sairaudet saattavat esimerkiksi vaikuttaa alkoholin käyttöön tai painoindeksiin. Geneettiset variantit, jotka ovat yhteydessä altisteeseen, eivät kuitenkaan riipu itse vasteen tasosta, jolloin käänteinen kausaalisuus pystytään sulkemaan pois.

**Ajallisen vaihtelun ja mittausvirheen minimointi.** Geneettiset variantit vaikuttavat altisteen pitkäaikaisiin tasoihin ja eivät ole niin alttiita suurille vaihteluille kuin yksittäiset mittaukset saattavat olla.

## 4 Aineisto ja tutkimuskysymykset

Tässä luvussa esitellään tutkielman aineistoa, sen sisältämiä altiste- ja vastemuuttujia sekä PRS-CS-menetelmällä luotuja instrumenttimuuttujia. Lisäksi selvitetään lyhyesti, miten analyysit tehtiin aiemmin esiteltyjä menetelmiä hyödyntäen.

### 4.1 Aineisto ja vastemuuttajat

**LASERI eli Lasten Sepelvaltimotaudin Riskitekijät -tutkimus.** Ensimmäisessä poikittaisaineistossa vuonna 1980 oli mukana yhteensä 3596 satunnaisesti valittua tyttöä ja poikaa. Heidän ikänsä olivat 3, 6, 9, 12, 15 tai 18 vuotta. Kohorttia seurattiin vuosina 1983, 1986, 1989, 2001 ja 2007 ja osallistujilta tutkittiin muun muassa verenpaine, kolesteroliarvot, painoindeksi sekä tupakanpolto [4]. Tähän tutkielmaan liittyvä seurantakäynti tapahtui vuonna 2011. Seurantatutkimuksessa vuonna 2011 mitattiin 2026 osallistujan kognitiivista toimintaa. [3]

Kognitiivisen toiminnan mittaamiseen käytettiin tietokonepohjaista testipatteristoa nimeltään CANTAB (the Cambridge Neuropsychological Test Automated Battery). CANTAB on tietokoneistettu, lähinnä ei-kielellinen ja kulttuurisesti neutraali laaja-alainen kognitiivinen testi. Testi tehdään käyttäen validoitua kosketusnäytöllistä tietokonetta. Täysi testipatteristo sisältää 24 yksittäistä testiä, joista voidaan valita sopivat tutkimuksesta riippuen. LASERI-tutkimuksessa testipatteristo valittiin siten, että sen voi suorittaa 20–30 minuutissa ja testit olivat sensitiivisiä ikääntymiselle. Nämä testit mittasivat muun muassa lyhytaikaista muistia, spatiaalista työmuistia, ongelmanratkaisukykyä, reaktioaikaa, huomiokykyä, nopeaa visuaalista prosessointia, visuaalista muistia, episodista muistia ja visuospatiaalista oppimista. [16]

Testikokonaisuus sisälsi 5 eri osatestiä:

1. Motor screening test (MOT), jossa tarkasteltiin yksilöiden testin suorittamiseen liittyviä vaikeuksia. Testin tarkoituksena oli tutustuttaa osallistuja testilaitteeseen sekä tunnistaa vaikeudet muun muassa näössä, liikkeissä ja ymmärryksessä.
2. Paired Associates Learning test (PAL), jossa tutkittiin visuaalista ja episodista muistia sekä visuospatiaalista oppimista.
3. Spatial Working Memory test (SWM), jonka avulla tutkittiin lyhytkestoista työmuistia, ongelmaratkaisukykyä sekä kykyä luoda itseorganisoitu strategia.
4. Reaction Time test (RTI), jossa mitattiin reaktionopeutta, liikenopeutta sekä huomiokykyä.
5. Rapid Visual Information Processing test (RVP), jossa tutkittiin visuaalisen informaation prosessointinopeutta, kykyä tunnistaa visuaalinen stimulus sekä pitkäkestoista huomiokykyä. [3] [16]

Kognitiivisista testeistä saatuun aineistoon oli sovellettu pääkomponenttianalyysiä kokoamaan yhteen jokaisen yksittäisen testin tulosten informaatio [16]. Näitä pääkomponenttianalyysillä muodostettuja muuttujia käytettiin myös tässä tutkielmas-  
sa vastemuuttujina. Aineistosta muodostettiin myös yleistä kognitiivista toimintaa

kuvaava muuttuja, jossa koottiin koko testipatteriston tuoma informaatio yhteen vastemuuttujaan. Tätä muuttujaa kutsutaan analyysissä kognitiiviseksi kokonaissummaksi. Aikaisemmissa tällä tutkimusaineistolla tehdyissä tutkimuksissa oli havaittu robusteja eli vakaita yhteyksiä kardiovaskulaaristen riskitekijöiden ja erityisesti muistia ja oppimista mittaavan PAL-testin välillä. Samanlaisia yhteyksiä oli havaittu myös eläinkokeissa. [3].

## 4.2 Altistemuuttujat

Altistemuuttujina tässä työssä käytetään systolista ja diastolista verenpainetta, seerumin *low density lipoprotein* (LDL)- ja *non-high density lipoprotein* (non-HDL)-kolesteroleja sekä painoindeksiä (*body mass index*, BMI). Systolinen verenpaine tarkoittaa verisuonen seinämään kohdistuvaa painetta sydämen supistuessa, kun taas diastolinen verenpaine on vastaava paine sydämen lepovaiheen aikana. LDL-kolesteroli on ns. huono kolesteroli, joka kertyy valtimoiden seinämiin aiheuttaen ateroskleroosin eli valtimonkovettumataudin kehittymisen. HDL-kolesteroli eli ns. hyvä kolesteroli puolestaan kuljettaa kolesterolia pois valtimoiden seinämistä. Non-HDL-kolesteroli on laskennallinen suure, joka kuvaa seerumin yhteenlaskettua kolesterolitasoa, josta on poistettu HDL-kolesterolin osuus. Painoindeksi on laskennallinen suure, joka kuvaa painon (kg) suhdetta pituuden neliöön ( $m^2$ ). Jokaisen riskitekijän osalta oli aiemmin laskettu kumulatiivista altistusta kuvaava muuttuja käyttäen seurannassa mitattujen arvojen perusteella määritetyn splini-käyrän alle jäävää aluetta (*area under the curve*, AUC). Splini-käyrällä tarkoitetaan kontrollipisteiden avulla piirrettyä sileää käyrää. AUC-muuttujat oli laskettu siten, että riskitekijöiden mitaukset huomioidaan lapsuudesta varhaiseen aikuisuuteen ikävuosien 6–24 välillä. Tutkielman altistemuuttujat valittiin aiemman tutkimuksen [3] sekä kiinnostavuuden perusteella.

## 4.3 Polygeeniset riskisummat

Tässä tutkielmassa instrumenttimuuttujina käytetyt altistekohtaiset polygeeniset riskisummat oli valmiiksi estimoitu käyttäen PRS-CS-menetelmää (*polygenic risk scores with continuous shrinkage*). PRS-CS on polygeeninen eli useaan geeniin perustuva ennustamismenetelmä, joka päättelee snipin posteriorivaikutuksen koon käyttäen genomilaaajuisia assosiaatioyhteenvedoja sekä ulkoisia kytkentäepätasapainon viitepaneeleja. PRS-CS hyödyntää moniulotteista bayesiläistä regressiota, jossa snippien vaikutuskoolle asetetaan jatkuvasti kutistuva prior. [17]

Aikaisemmin monet sovellukset laskivat PRS:n geneettisten markkereiden osajoukosta sen jälkeen, kun kytkentäepätasapainoiset snipit oli karsittu ja myös testattu p-arvon kynnsarvolla GWAS-analyysissä. Vaikka tällä menetelmällä oli hyötyjä sekä laskennallisuuden että käsitteellisen yksinkertaisuuden osalta ja sitä käytettiin ennustamaan geneettistä alttiutta laajan fenotyypin kirjon osalta, tutkimukset ovat osoittaneet, että tämä tavanomainen menetelmä polygeenisen riskisumman muodostamiseksi hylkää informaatiota ja rajoittaa ennustetarkkuutta. [17]

## 4.4 Aineiston koko ja mallien implementointi

Kaikki analyysissä käytettävät muuttujat ovat jatkuvia. Kunkin osa-analyysin polygeeninen riskisumma, eli instrumenttimuuttuja  $Z$ , kardiovaskulaarinen riski- eli altistemuuttuja  $X$  sekä kognitiivista toimintaa kuvaava vastemuuttuja  $Y$ , on jatkuva. Tästä johtuen ei voida käyttää Hernánin [9] esimerkin mukaista mallia (2), jossa kaikki muuttujat ovat dikotomisista 1/0-muuttujia, vaan käytetään mallia (11). Polygeeniset riskisummat oli laskettu 1815 tutkittavalta ja riskitekijäaltistusta kuvaavat AUC-muuttujat 2026 tutkittavalta. Osatestistä riippuen kognitiivista toimintaa määrittävät vastemuuttujat oli määritetty 1822–2011 tutkittavalta. Testikohtaiset erot tutkittavien lukumäärässä johtuvat testitilanteen teknisistä ongelmista tai tutkittavien kieltäytymisestä tehdä yksittäisiä testiosioita. Yhteensä 1798 tutkittavaa on tehnyt kaikki neljä kognitiivisen toiminnan testin osatestiä (seulontatestin lisäksi) [16]. Kaikki tutkielmassa käytetyt altiste- ja vastemuuttujat on esitetty taulukossa 1. Jokaiselle viidelle yksittäiselle altisteelle tehtiin erikseen oma osa-analyysinsä jokaisen viiden kognitiivisen vastemuuttujan osalta. Yksittäisiä analyysejä on siten kokonaisuudessaan  $5 \times 5 = 25$ . Yksilömäärien vaihteluväli näissä osa-analyyseissä oli 1609–1800.

Altistemuuttujat	Vastemuuttujat
Systolinen verenpaine	PAL-testi
Diastolinen verenpaine	SWM-testi
Non-HDL-kolesteroli	RTI-testi
LDL-kolesteroli	RVP-testi
Painoindeksi	Kognitiivinen kokonaissumma

Taulukko 1: Tutkielmassa käytetyt altiste- ja vastemuuttujat. Jokaiselle altistemuuttujalle oli lisäksi oma polygeeninen riskisummansa eli instrumenttimuuttujansa.

Analyysissä käytettiin sekä kaksivaiheista pienimmän neliösumman menetelmää että kolmiomenetelmää (ks. luku 3.2). Kaksivaiheisen pienimmän neliösumman menetelmä toteutettiin sekä käsin ohjelmoiden että R:n valmiilla `tsls`-funktiolla paketista `sem` [18]. Kolmiomenetelmä toteutettiin painotetulla käänteisen varianssin menetelmällä, jossa painot eivät tässä tapauksessa kuitenkaan vaikuta, koska polygeeninen riskisumma on ainoana instrumenttimuuttujana. Painotetun käänteisen varianssin menetelmä toteutettiin R:ssä paketin `MendelianRandomization` funktiolla `mr_ivw` [19], joka raportoi kausaalisen vaikutuksen estimaatin ja sen keskivirheen sekä normaalijakaumaan perustuvan 95 prosentin luottamusvälin. Jos instrumenttimuuttujana käytettäisiin useampaa yksittäistä geneettistä varianttia, funktio suorittaisi myös heterogeisuus-testin, joka kertoo, ovatko yksittäisten varianttien kausaalisen vaikutuksen estimaatit keskenään samanlaisia vai eivät [15]. Kolmiomenetelmän piste-estimaatin oikeellisuuden varmistamiseksi kolmiomenetelmä laskettiin myös käsin R:ssä tosin vain piste-estimaatin osalta.

Tutkielman aineiston muokkaamiseen käytettiin SAS-ohjelman online-versiota. Analyysien tekemiseen ja kuvien muodostamiseen käytettiin R Studiota ja R:n versiota 4.2.1. Käytetyt ohjelmakoodit sekä tulosteet ovat liitteissä tutkielman lopussa. SAS-koodi on liitteessä A ja R-koodi sekä -tuloste liitteessä B.

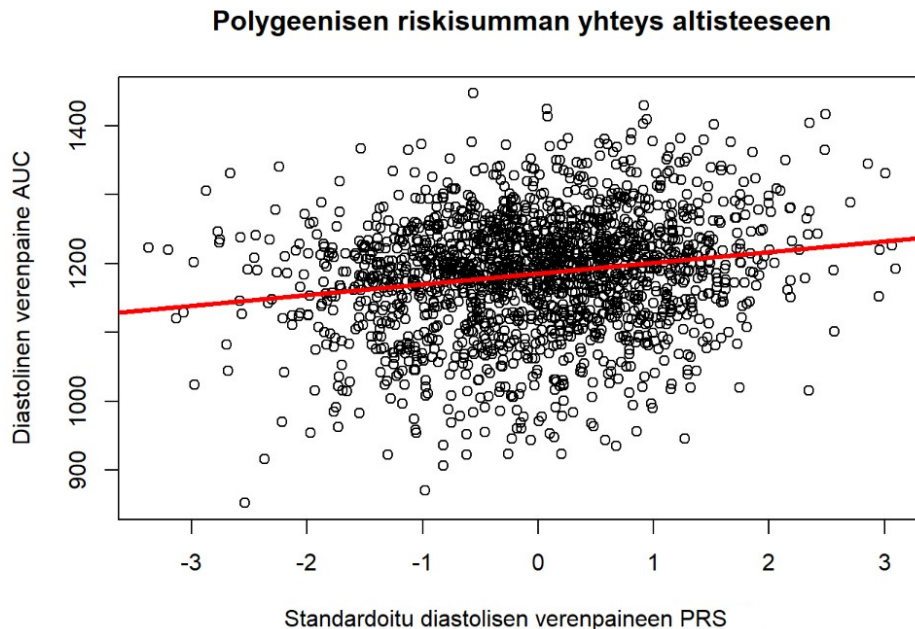
## 5 Tulokset

Tarkastellaan mendelistisellä satunnaistamisella saatuja eri riskitekijöiden kausaalisen vaikutuksen estimaatteja erityisesti PAL-testiin liittyen. Tarkemmat analyysin vaiheet sekä kaikki tulokset löytyvät liitteestä B.

### 5.1 Altisteen ja instrumenttimuuttujan välinen yhteys

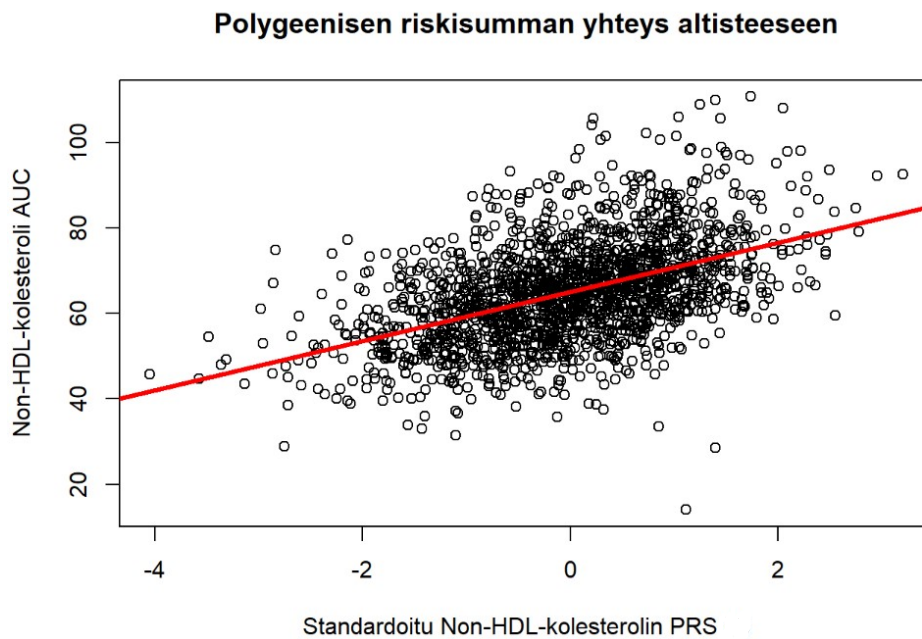
Tarkastellaan aluksi polygeenisten riskisummien ja altistemuuttujien välistä yhteyttä. Instrumenttimuuttujan yhtenä ehtona oli, että sillä on oltava yhteys altistemuuttujan kanssa, jotta se on ylipäättään kelvollinen instrumenttimuuttujaksi. Yhteys todennettiin lineaarisen regressiomallin kuvaajan, muuttujien välisen Pearson-korrelaation ja F-testisuuren arvon avulla. Pearsonin korrelaatiokerroin mittaa lineaarisen yhteyden voimakkuutta kahden muuttujan välillä.

Kuvasta 3 nähdään diastolisen verenpaineen ja sen oman instrumenttimuuttujan välinen yhteys. Muuttujien välinen korrelaatiokerroin oli 0.18 eli muuttujien välillä on vähintäänkin heikko lineaarinen yhteys. Instrumenttimuuttujan ehdoissa korrelaation piti poiketa nolasta, joten ehdon katsotaan siten täyttyvän. Systolisen verenpaineen yhteys sen instrumenttimuuttujaan oli hyvin samankaltainen korrelaatiokertoimen arvon ollessa 0.23 eli hieman suurempi. F-testisuuren arvon tuli olla vähintään 10 instrumenttimuuttujan ja altisteen välisessä regressiomallissa. Diastolisen verenpaineen mallin F-arvo oli 62.4 ja systolisen verenpaineen 100.5 eli F-testisuuren arvon ehdon voidaan nähdä täyttyvän.



Kuva 3: Diastolisen verenpaineen lineaarinen yhteys sen polygeeniseen riskisummaan (PRS).

Kolesterolien osalta korrelaatio oli non-HDL-kolesterolin tapauksessa noin 0.49 (kuva 4) kuten myös LDL-kolesterolin tapauksessa eli korrelaatio on selvästi vahvem-



Kuva 4: Non-HDL-kolesterolin lineaarinen yhteys sen polygeeniseen riskisummaan (PRS).

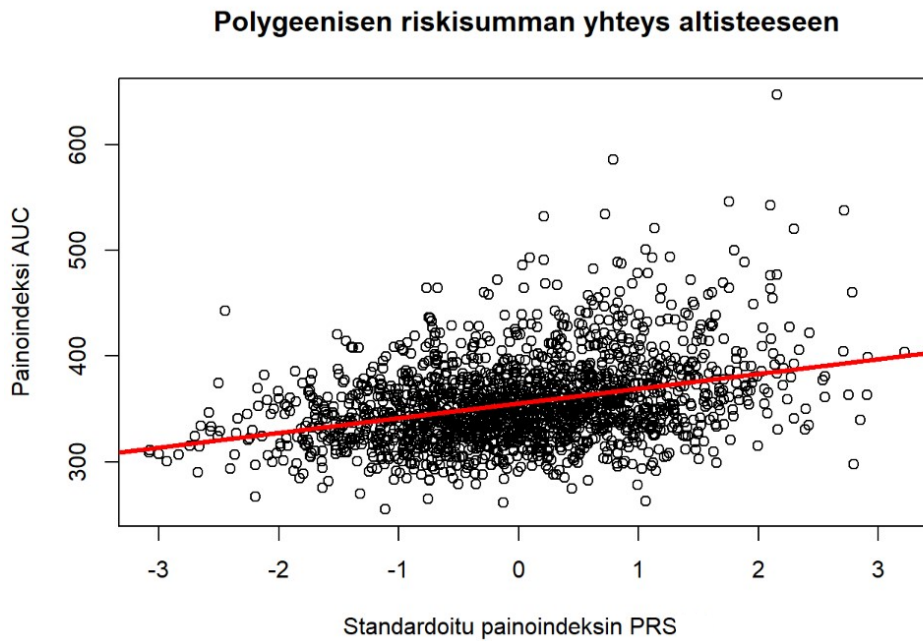
paa altisteen ja polygeenisen riskisumman välillä verrattuna verenpaineen vastaaviin arvoihin. Se, että korrelaatiokerroin ei juurikaan poikkea LDL:n ja non-HDL:n välillä, selittyy sillä, että LDL- ja non-HDL-arvot ovat useimmilla yksilöillä lähes samat. F-testisuureen arvot olivat non-HDL:llä 558.4 ja LDL:llä 580.8 eli hyvin suuret.

Painoindeksi eli BMI:n kohdalla Pearsonin korrelaatiokerroin oli 0.34 (kuva 5) eli yhteyden vahvuus on verenpaineen ja kolesterolin omien yhteyksien vahvuuksien väliltä. F-testisuureen arvo oli myös edellisten väliltä sen ollessa 241.6.

Kaikki altistemuuttajat olivat korreloituneita omien instrumenttimuuttajiensa kanssa, joten kaikki instrumenttimuuttajina toimivat polygeeniset riskisummat vaikuttavat valideilta. Jokaisen riskisumman yhteyden vahvuus omaan altisteseensa on esitetty taulukossa 2. Seuraavaksi voidaan siirtyä itse kausaalisuutta koskevien analyysien pariin.

Instrumenttimuuttajien vahvuus			
Altistemuuttaja	Piste-estimaatti (vaikutus)	Korrelaatio (Pearson)	F-arvo
Systolinen verenpaine	28.361	0.23	100.5
Diastolinen verenpaine	16.635	0.18	62.4
Non-HDL-kolesteroli	5.767	0.49	558.4
LDL-kolesteroli	5.673	0.49	580.8
Painoindeksi	13.897	0.34	241.6

Taulukko 2: Polygeenisten riskisummien yhteys altistemuuttajiin. Jokainen instrumentti-altiste-yhteys näyttää tarpeeksi vahvalta.



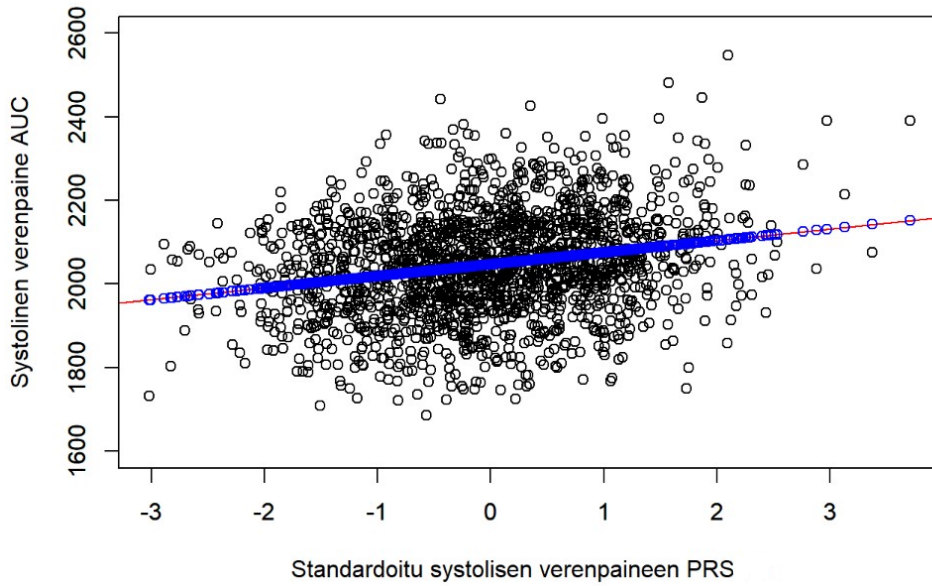
Kuva 5: Painoindeksiin lineaarinen yhteys sen polygeeniseen riskisummaan (PRS).

## 5.2 Altisteen ja vasteen välinen kausaalinen yhteys

Tässä luvussa tarkastellaan altistemuuttujien eli kardiovaskulaaristen riskitekijöiden ja vastemuuttujien eli kognitiivisen toiminnan välistä kausaalista yhteyttä mendelisisestä satunnaistamisesta saatavilla tuloksilla. Vertailukohtana tuloksille käytetään perinteistä lineaarista regressiomallia, jossa populaation yksilöiden vasteiden ainoana selittävänä tekijänä on yksilöiden altisteen havaitut arvot suoraan aineistosta. Tätä vertailuna käytettävää (mahdollisesti sekoittunutta) mallia kutsutaan tästä eteenpäin alkuperäiseksi malliksi, vaikka alkuperäisessä Rovion ym. tutkimuksessa [3] myös kardiovaskulaaristen riskitekijöiden muuttujat oli standardoitu. Tuloksia esitetään tässä pääasiassa PAL-testiin liittyen, mutta myös muita mielenkiintoisia tuloksia on pyritty nostamaan esiin. Loput analyysitulosteet löytyvät liitteenä tutkielman lopusta.

### 5.2.1 Verenpaine

Kaksivaiheisen pienimmän neliösumman menetelmän ensimmäisessä vaiheessa altisteelle ennustettiin arvot instrumenttimuuttujan avulla. Systolisen verenpaineen ennustetut arvot ovat nähtävissä kuvasta 6, jossa siniset ympyrät merkitsevät seuraavassa vaiheessa käytettäviä ennustettuja altisteen arvoja yksilöille. Toisen vaiheen regressiomallista saatavan vaikutuksen estimaatin 95 prosentin luottamusväli  $[-0.0017, 0.0017]$  sisältää nollan, mikä tulkitaan niin, ettei systolisella verenpaineella ole kausaalista vaikutusta PAL-testin tulokseen. Kausaalisen vaikutuksen p-arvo on lähes 0.99, mikä tarkoittaa, että jopa noin 99 prosentin todennäköisyydellä toistetussa aineistonkeruussa saadaan tätä aineistoa äärevämpi aineisto, kun nollahypoteesi pitää paikkansa eli kausaalista vaikutusta ei ole. Alkuperäisillä altisteen arvoilla tehdystä regressiomallista saatavan vaikutuksen estimaatin 95 prosentin luottamusväli



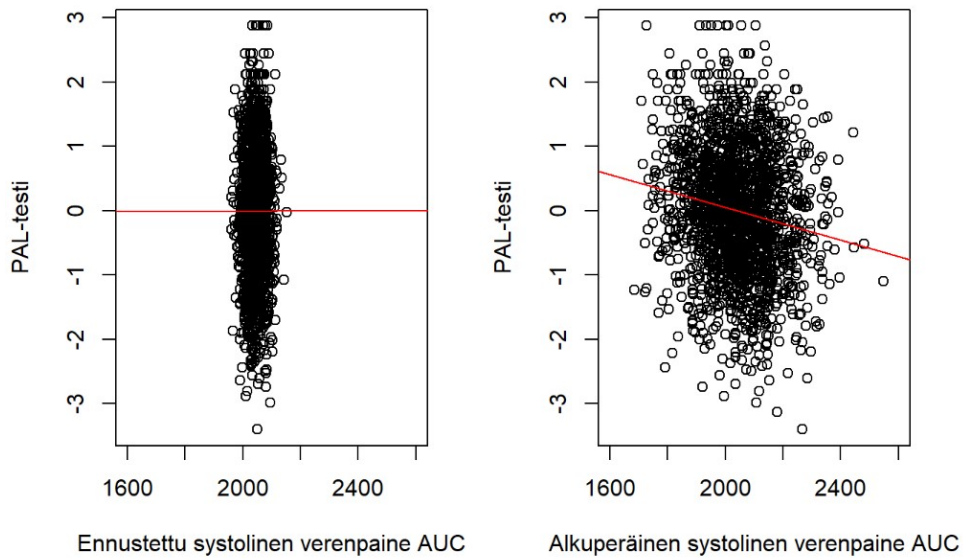
Kuva 6: Systolisen verenpaineen yhteys sen polygeeniseen riskisummaan (PRS). Regressiosuoralla olevat siniset ympyrät kuvaavat yksilöiden sovitettuja arvoja.

$[-0.0016, -0.0009]$  ei sisällä nollaa, mikä tarkoittaa, että systolisella verenpaineella olisi vaikutus PAL-testin tulokseen. Alkuperäisen mallin p-arvokin on hyvin pieni  $1.74 * 10^{-11}$ .

Kaksivaiheisen pienimmän neliösumman mallin estimoima systolisen verenpaineen kausaalinen vaikutus sekä suoraan altisteen havaitusta arvosta estimoitu systolisen verenpaineen vaikutus ovat nähtävissä kuvasta 7. Koska molempien mallien kuvaajien vaaka-akselin arvot on esitetty samalla skaalalla, osuvat 2SLS-mallin altisteen arvot vasemmanpuoleisessa kuvaajassa hyvin lähekkäin toisiaan. Tämä johtuu siitä, että arvot ovat polygeenisten riskisummien avulla ennustettuja. Tämä voidaan mallin idean perusteella ajatella olevan niin, että yksilöiden altisteen arvot ovat nyt sillä tasolla, millä ne pelkästään geenien perusteella olisivat, jos ympäristöllä ei olisi niihin vaikutusta. Jatkossa myös 2SLS-mallin ennustetut arvot ovat tasaisemmin kuvaajan leveydellä eli  $x$ -akselin arvot ovat pienemmällä välillä verrattuna alkuperäisen mallin  $x$ -arvoihin. Tästä suurennuksesta johtuen kausaalisen vaikutuksen regressiosuora hieman loivenee kohti nollaa, jonka takia onkin tärkeää tehdä tarkemmat päätelmät piste-estimaattien ja luottamusvälien perusteella.

Kolmiomenetelmästä (IVW) saatavan systolisen verenpaineen kausaalisen vaikutuksen estimaatin 95 prosentin luottamusväli on  $[-0.002, 0.002]$  eli se sisältää myös nollan. P-arvo on myös kaksivaiheisen pienimmän neliösumman menetelmän tapaan lähes 0.99. Edellä mainituilla menetelmillä estimoidut systolisen verenpaineen vaikutuksen piste-estimaatit, luottamusvälit ja p-arvot PAL-testiin liittyen on esitetty taulukossa 3.

Eri menetelmät antoivat kaikissa analyyseissä poikkeuksetta samoja tuloksia, joten eri menetelmien tulosten yhtäläisyyttä ei jatkossa nosteta erikseen esiin, ellei siihen ole erityistä tai poikkeavaa syytä. Luottamusvälin oletetaan myös olevan aina piste-estimaatin 95 prosentin luottamusväli, ellei toisin mainita.



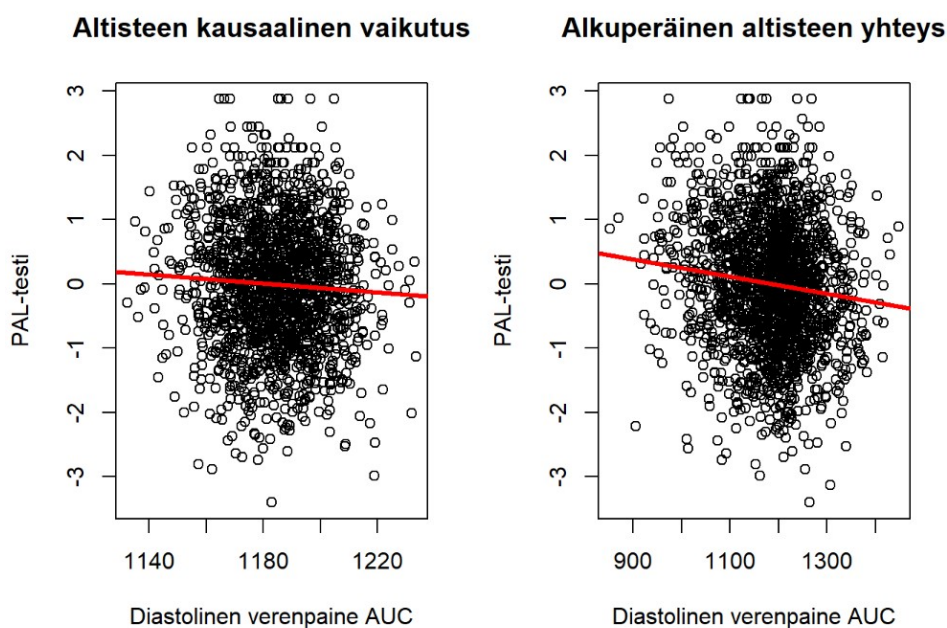
Kuva 7: Vasemmalla 2SLS-menetelmällä estimoitu systolisen verenpaineen kausaalinen vaikutus ja oikealla suoraan altistemuuttujan eli systolisen verenpaineen AUC-muuttujan havaitusta arvosta estimoitu systolisen verenpaineen vaikutus PAL-testin arvoon.

Systolinen verenpaine – PAL-testi			
Menetelmä	Piste-estimaatti (vaikutus)	Luottamusväli (95%)	P-arvo
2SLS	$1.33 * 10^{-5}$	-0.002, 0.002	0.988
IVW	$1.34 * 10^{-5}$	-0.002, 0.002	0.988
LM	$-1.27 * 10^{-3}$	-0.002, -0.001	$1.74 * 10^{-11}$

Taulukko 3: Systolisen verenpaineen yhteys PAL-testiin. LM viittaa vertailukohteena olevaan alkuperäiseen lineaariseen malliin.

Diastolisen verenpaineen kausaalisen vaikutuksen luottamusväli 2SLS-mallissa on  $[-0.0064, -0.0004]$  (p-arvo 0.03) ja kolmiomenetelmässä, jossa tuloste ilmoittaa luvut vain kolmen desimaalin tarkkuudella,  $[-0.007, -0.000]$  (p-arvo 0.03) eli molempien mukaan korkeammalla diastolisella verenpaineella on heikentävä kausaalinen vaikutus PAL-testin tulokseen. Alkuperäisen mallin luottamusväli vaikutukselle on  $[-0.0019, -0.0008]$  eli vaikutuksen olemassaolon tulkinta on samansuuntainen. Alkuperäisen mallin p-arvo on vielä selvästi äärevämpi sen ollessa  $5.71 * 10^{-7}$ . Diastolisen verenpaineen vaikutusten kuvaajat on esitetty kuvassa 8 ja tulokset taulukossa 4 PAL-testiin liittyen.

Korkeammalla diastolisella verenpaineella näyttäisi olevan heikentävä kausaalinen vaikutus myös RVP- eli visuaalisen prosessoinnin testissä suoriutumiseen (liite B, s. 31–32). Alkuperäisessä mallissa diastolisella verenpaineella ja RVP-testillä ei ollut merkittävää yhteyttä. Diastolisella verenpaineella näytti myös olevan heikentävä kausaalinen vaikutus kaikki kognitiiviset testit yhteen koostavan kognitiivisen kokonaissummatestin tuloksiin (liite B, s. 33–34). Alkuperäisessä mallissa tämä



Kuva 8: Vasemmalla 2SLS-menetelmällä estimoitu diastolisen verenpaineen kausaalinen vaikutus ja oikealla suoraan altistemuuttujan eli diastolisen verenpaineen AUC-muuttujan havaitusta arvosta estimoitu diastolisen verenpaineen vaikutus PAL-testin arvoon. Molempien mallien mukaan korkeammalla diastolisella verenpaineella on heikentävä vaikutus PAL-testin tulokseen.

Diastolinen verenpaine – PAL-testi			
Menetelmä	Piste-estimaatti (vaikutus)	Luottamusväli (95%)	P-arvo
2SLS	-0.0034	-0.006, -0.000	0.028
IVW	-0.0035	-0.007, -0.000	0.027
LM	-0.0013	-0.002, -0.001	$5.71 \times 10^{-7}$

Taulukko 4: Diastolisen verenpaineen yhteys PAL-testiin. Kaikki menetelmät antoivat tilastollisesti merkitsevän tuloksen.

altiste-vaste-yhteys oli tosin samaan suuntaan vielä selvästi vahvempi.

### 5.2.2 Kolesterolit

Kolesterolin osalta 2SLS-mallin luottamusvälit olivat non-HDL:llä [-0.0043, 0.0127] (p-arvo 0.33) ja LDL:llä [-0.0058, 0.0114] (p-arvo 0.52). Kolmiomenetelmästä saadut luottamusvälit olivat non-HDL:llä [-0.004, 0.012] (p-arvo 0.33) ja LDL:llä [-0.006, 0.011] (p-arvo 0.52). Alkuperäisen mallin luottamusvälit olivat non-HDL:llä [-0.0116, -0.004] (p-arvo  $6.12 \times 10^{-5}$ ) ja LDL:llä [-0.0146, -0.0068] (p-arvo  $7.86 \times 10^{-8}$ ). Kolesterolilla ei siis vaikuttaisi olevan kausaalista vaikutusta PAL-testin tulokseen, vaikka yhteys alkuperäisen mallin perusteella näyttäisikin olevan. Nämä tulokset on koottu taulukkoon 5.

Non-HDL-kolesteroli – PAL-testi			
Menetelmä	Piste-estimaatti (vaikutus)	Luottamusväli (95%)	P-arvo
2SLS	0.0042	-0.004, 0.013	0.328
IVW	0.0042	-0.004, 0.012	0.325
LM	-0.0078	-0.012, -0.004	$6.12 * 10^{-5}$
LDL-kolesteroli – PAL-testi			
Menetelmä	Piste-estimaatti (vaikutus)	Luottamusväli (95%)	P-arvo
2SLS	0.0028	-0.006, 0.011	0.524
IVW	0.0028	-0.006, 0.011	0.522
LM	-0.0107	-0.015, -0.007	$7.86 * 10^{-8}$

Taulukko 5: Kolesterolien yhteys PAL-testiin. Mendelistisellä satunnaistamisella ei pystytty todentamaan lineaarisella mallilla havaittavaa yhteyttä.

### 5.2.3 Painoindeksi

Painoindeksillä ei ollut vaikutusta PAL-testin tulokseen mendelistisen satunnaistamisen mallilla eikä alkuperäisellä mallilla, joten se sivuutetaan tässä kohtaa. Sen sijaan mielenkiintoista oli se, että reaktio- eli RTI-testissä suuremmalla painoindeksillä oli heikentävä kausaalinen vaikutus testin tulokseen (liite B, s. 59–60). Alkuperäisellä mallilla ei ollut merkitsevää yhteyttä painoindeksin ja RTI-testin tuloksen välillä. Mielenkiintoista oli myös se, että painoindeksillä oli heikentävä kausaalinen vaikutus RVP-testin tuloksiin (liite B, s. 61–62). Tässäkään tapauksessa merkitsevää yhteyttä alkuperäisten altisteen arvojen ja RVP-testin tulosten välillä ei ollut. Aikaisemmissa LASERI-tutkimuksissa oli kuitenkin havaittu painoindeksin merkitsevä yhteys RVP-testin tuloksiin vertaamalla ryhmää, jolla on ollut korkea BMI läpi eliniän, ryhmään, jolla BMI on ollut normaalilla tasolla [20]. Nyt suuremman painoindeksin kausaalisen heikentävän vaikutuksen RVP-testiin voidaan katsoa koskevan koko populaatiota, kun vaikutuksen oletetaan olevan lineaarinen. Painoindeksin tulokset RTI- ja RVP-testiin liittyen on esitetty taulukossa 6.

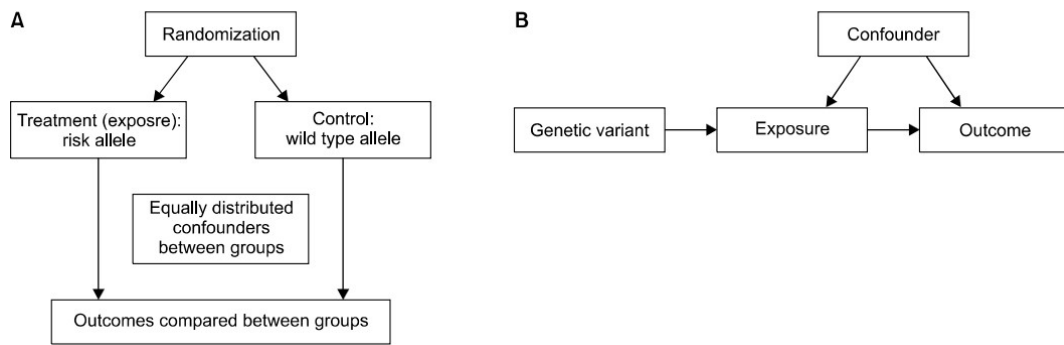
Painoindeksi – RTI-testi			
Menetelmä	Piste-estimaatti (vaikutus)	Luottamusväli (95%)	P-arvo
2SLS	-0.0047	-0.008, -0.001	0.012
IVW	-0.0046	-0.008, -0.001	0.010
LM	0.0009	-0.000, 0.002	0.117
Painoindeksi – RVP-testi			
Menetelmä	Piste-estimaatti (vaikutus)	Luottamusväli (95%)	P-arvo
2SLS	-0.0045	-0.008, -0.001	0.009
IVW	-0.0045	-0.008, -0.001	0.008
LM	-0.0011	-0.002, 0.000	0.058

Taulukko 6: Painoindeksin yhteys RTI- ja RVP-testiin. Korkeammalla painoindeksillä näyttäisi olevan kausaalisesti heikentävä vaikutus kummassakin testissä suoritukseen. Alkuperäisellä mallilla tilastollisesti merkitsevää yhteyttä ei havaita ( $p > 0.05$ ).

## 6 Pohdinta

Instrumenttimuuttujan tarkoituksena on minimoida sekoittavien tekijöiden vaikutus ja siten auttaa estimoimaan altisteen ja vasteen välinen kausaalinen yhteys havainnoivassa tutkimuksessa. Tässä työssä instrumenttimuuttujana toimivat geneettisen variaation perusteella muodostetut polygeeniset riskisummat eli kyseessä on instrumenttimuuttujan erikoistapaus mendelistinen satunnaistaminen eli MR. Menetelmässä pyritään geenitiedon avulla määrittämään altisteen pitkäaikaisempi taso ja sulkemaan pois altisteeseen ja vasteeseen vaikuttavat sekoittavat tekijät. Esimerkiksi alkoholin käyttö saattaisi vaikuttaa sekä verenpaineeseen että kognitiivisiin toimintoihin. Tällöin verenpaineen vaikutus kognitiivisiin toimintoihin saattaisi vaikuttaa suuremmalta kuin se todellisuudessa on, jos alkoholilla on yhteys sekä verenpaineeseen että kognitiivisiin toimintoihin. Geenitiedon avulla pystytään estimoimaan altisteen tasot ilman ympäristön muokkaamaa vaikutusta, jolloin tavoitteena on, että instrumenttimuuttuja korjaa pelkän verenpaineen vaikutuksen kognitiivisiin toimintoihin liittyen minimoimalla sekoittavien tekijöiden vaikutuksen. Kuten useaan tilastolliseen menetelmään, myös mendelistiseen satunnaistamiseen liittyy monia oletuksia sekä epävarmuutta.

Mendelistisen satunnaistamisen idean voidaan katsoa olevan muuttaa havainnoiva tutkimus vastaamaan satunnaistettua kontrolloitua koetta. Tämä ajatus toteutuu, koska alleelien (geenien) satunnainen jakautuminen erottaa yksilöt altistettuihin ja kontrolloituihin ryhmiin, mikä johtaa siihen, että sekoittavat tekijät jakautuvat tasaisesti ryhmien välillä. Tämä idea on nähtävissä kuvasta 9. Satunnaistamisen idean voidaan näin katsoa täyttyvän, jolloin isoimmat mendelistiseen satunnaistamiseen liittyvistä oletuksista ja epävarmuuksista liittyvätkin instrumenttimuuttujalta vaadittuihin ehtoihin. Ensimmäisenä ehtona oli geneettisen variantin eli tämän tutkielman tapauksessa polygeenisen riskisumman yhteys altisteeseen. Polygeeniset riskisummat oli tehty valmiiksi PRS-CS-menetelmää käyttäen yhdistämällä valta-



Kuva 9: (A) Satunnaistetun kokeellisen kokeen ja mendelistisen satunnaistamismenetelmän yhteys. Yksilöt jakautuvat itsestään altiste- ja kontrolliryhmiin satunnaisesti perittyjen geenien perusteella. (B) Mendelistisen satunnaistamisen perusajatus (ks. kuva 1). Kuvan lähde: [14]

va määrä sopivia snippejä, jotka ovat yhteydessä altisteeseen. Yhteys varmistettiin vielä tässä tutkielmassa tarkastelemalla lineaarista yhteyttä sekä korrelaatiokerrointa, joten ehdon voidaan katsoa täyttyvän. Seuraavien ehtojen täyttymisen takaaminen on hankalampaa. Toisena ehtona oli, että instrumenttimuuttuja saa vaikuttaa vastemuuttujaan ainoastaan altistemuuttujan kautta. Tätä ehtoa voidaan kutsua myös "ei pleiotropiaa -ehdoksi". Pleiotropialla tarkoitetaan tilannetta, jossa sama geeni eli tässä tapauksessa polygeeninen riskisumma vaikuttaisi useampaan yksilössä olevaan ominaisuuteen. Tässä tapauksessa geenit voisivat vaikuttaa siten myös mittaamattomien kovariaattien arvoihin, jolloin kausaalisen vaikutuksen estimointi johtaisi edelleen ainakin jossakin määrin harhaiseen estimaattiin ja päätelmään. MR-mallin eri laajennuksia, kuten MR-Egger-menetelmää tai painotettua mediaanitestiä, voidaan käyttää pleiotropian havaitsemiseen ja/tai altistuksen kausaalivaikutuksen arvioimiseen, vaikka oletus 2 ja/tai 3 rikottaisiin. Tässä tutkielmassa tätä ei sen tarkemmin tarkastella, vaan oletetaan, että useasta snipistä muodostetut polygeeniset riskisummat eivät ole pleiotrooppisia. Kolmas ehto liittyy instrumenttimuuttujan riippumattomuuteen sekoittavista tekijöistä. Jos instrumenttimuuttuja korreloi myös sekoittavien tekijöiden kanssa, on kausaalisen vaikutuksen estimaatti edelleen jonkin verran sekoittunut. Vaikka on teoreettisesti mahdotonta todistaa, ettei tämä tilanne olisi lainkaan läsnä MR-kokeessa, on tunnettujen sekoittavien tekijöiden ja geneettisen variantin välinen korrelaatio kuitenkin mahdollista testata ja kumota. [14]

## 6.1 Yhteenvedo ja johtopäätökset

Systolisella verenpaineella ei näyttäisi olevan kausaalista vaikutusta minkään keskiässä suoritettun kognitiivisen testin tulokseen, kun taas korkeammalla diastolisella verenpaineella näyttäisi olevan heikentävä kausaalinen yhteys muistiin ja oppimiseen (PAL-testi), informaatioprosessointiin (RVP-testi) sekä yleiseen kognitiiviseen toimintaan (koko CANTAB-testipatteri). Kolesteroliarvoilla ei näyttäisi olevan kausaalista vaikutusta kognitiivisiin toimintoihin. Korkeammalla painoindeksillä näyttäisi

si olevan heikentävä kausaalinen vaikutus reaktioaikaan (RTI-testi). Tämä oli myös mielenkiintoinen havainto sikäli, että alkuperäisen lineaarisen mallin regressiosuoran suunta oli niukasti toiseen suuntaan, vaikkakin merkitsevää tilastollista yhteyttä ei sen perusteella todeta. Korkeampi painoindeksi viittaa myös heikentävään kausaaliseen vaikutukseen informaatioprosessoinnin (RVP-testi) osalta.

Aikaisemmin Rovion ym. tutkimuksessa havaittuja riskitekijöiden ja kognitiivisten toimintojen välisiä yhteyksiä ei pystytty tässä tutkielmassa kaikilta osin osoittamaan kausaaliseksi [3]. On syytä kuitenkin ottaa huomioon se, että vaikka kaikkien instrumenttimuuttujia koskevien ehtojen oletettaisiin täyttyvän, voisi aineiston koko olla suurempi, jolloin tilastollista voimaa olisi myös enemmän. Tilastollisen voiman puute näkyy esimerkiksi siinä, että MR-menetelmällä saatavat luottamusvälit ovat leveämpiä kuin perinteisellä lineaarisella regressiolla estimoidut luottamusvälit. Tämä voi johtaa siihen, että vaikka vaikutuksen estimaatti olisi lähempänä nollaa alkuperäisessä mallissa kuin MR-menetelmällä, sisältää MR-menetelmän luottamusväli silti nollan ja alkuperäisen mallin ei. Tällöin tulkintana on se, ettei kausaalista vaikutusta ole, vaikka alkuperäisen mallin tulkintana on, että vaikutus olisi olemassa. Tämä johtuu yksinkertaisesti siitä, että MR-menetelmällä saadut keskivirheet ovat suurempia, jolloin myös luottamusvälit ovat leveämpiä, vaikka molemmat mallit on estimoitu samalla otoskoolla. Suuremmalla otoskoolla saataisiin siten vielä tarkempia tuloksia, jolloin myös päätelmät olisivat luotettavampia.

MR-menetelmien luotettavuuteen on silti tukeuduttava, sillä MR-tutkimukset ovat samantapaisia kuin yleisesti suositellut satunnaistetut kontrolloidut kokeet silloin, kun kolme keskeistä instrumenttimuuttujan oletusta täyttyvät. Vaikka MR-menetelmä ei korvaa kontrolloitua tutkimusta, se on silti hyödyllinen tilanteissa, joissa kontrolloitu tutkimus saattaisi olla epäeettinen tai mahdoton toteuttaa. MR-menetelmä voi tarjota tarkempia arvioita riskitekijöiden kausaalivaikutuksesta vastemuuttujaan eli tässä tapauksessa kognitiivisiin toimintoihin verrattuna havainnoiviin tutkimuksiin, sillä se pystyy ylittämään niiden rajoituksia. Mendelistisen satunnaistamisen lähestymistapaa käytetään yhä enemmän arvioimaan kausaalisuutta riskitekijöiden ja sairauksien välillä. Tutkimusten tulosten luotettavuus riippuu kuitenkin useiden oletusten oikeellisuudesta, jotka on tarkastettava huolellisesti ja tulkittava aiemman biologisen tiedon kontekstissa. Hyvin toteutetut MR-tutkimukset voivat olla arvokas menetelmä kausaalisuhteiden tutkimiseen monimutkaisten sairauksien tapauksissa. [14]

## 6.2 Jatkotutkimuksen suuntia

Tulevaisuudessa aiheesta voisi tehdä vielä perusteellisemmän analyysin eri riskitekijöitä käyttäen. Malleihin voisi myös lisätä kovariaatteja tai vähintäänkin muuttujat sukupuolesta ja iästä kuten oli tehty osassa aikaisemman tutkimuksen analyyseissä [3]. Tämän tutkielman analyyseissä malleihin ei lisätty muita muuttujia eikä tietoa sukupuolesta tai tarkemmasta iästä, vaan kaikkien riskitekijöiden arvot olivat 6 ja 24 ikävuoden kokonaisajalta. Tarkemmin alaa tuntevien ja lääketieteellisiä yhteyksiä paremmin ymmärtävien asiantuntijoiden on mahdollista myös pohtia ja tutkia, mitkä sekoittavat tekijät saattaisivat aiheuttaa aikaisemmissä analyyseissä todetun yhteyden riskitekijöiden ja kognitiivisten toimintojen välillä, jos itse kausaalivaiku-

tusta ei näytetty mendelistisellä satunnaistamisella olevan.

Polygeenisten riskisummien pleiotropiaa on myös mahdollista tutkia tarkemmin käyttämällä esimerkiksi MR Egger -menetelmää kausaalisen vaikutuksen estimoimiseksi, jos on syytä epäillä pleiotropian esiintymistä. Tutkielman lopusta löytyvät tässä tutkielmassa käytetyt R-koodit jatkoanalyysien tekoon ovat vapaasti tutkijoiden ja muidenkin käytettävissä mihin tahansa hyötytarkoitukseen.

## Viitteet

- [1] Teumer, A. (2018). *Common Methods for Performing Mendelian Randomization*. *Frontiers in Cardiovascular Medicine*, 5(51).
- [2] Burgess, S., Davey Smith, G., Davies, M., Dudbridge, F., Gill, D., Glymour, M., Hartwig, F., Holmes, M., Minelli, C., Relton, C., Theodoratou, E. (2020). *Guidelines for performing Mendelian randomization investigations [version 2; peer review: 2 approved]*. Wellcome Open Research, 4.
- [3] Rovio, S., Pahkala, K., Nevalainen, J., Juonala, M., Salo, P., Kähönen, M., Hutri-Kähönen, N., Lehtimäki, T., Jokinen, E., Laitinen, T., Taittonen, L., Tossavainen, P., Viikari, J., MD, Rinne, J., Raitakari, O. (2017). *Cardiovascular Risk Factors From Childhood and Midlife Cognitive Performance*. *Journal Of The American College Of Cardiology*, 69(18), 2279–2289.
- [4] Raitakari, O., Juonala, M., Rönnemaa, T., Keltikangas-Järvinen, L., Räsänen, L., Pietikäinen, M., Hutri-Kähönen, N., Taittonen, L., Jokinen, E., Marniemi, J., Jula, A., Telama, R., Kähönen, M., Lehtimäki, T., Akerblom, H., Viikari, J. (2008). *Cohort Profile: The Cardiovascular Risk in Young Finns Study*. *International Journal of Epidemiology*, 37(6), 1220–1226. PMID: 18263651.
- [5] Burgess, S., Small, D., Thompson S. (2017). *A Review of Instrumental Variable Estimators for Mendelian Randomization*. *Statistical Methods in Medical Research*, 26(5), 2333–2355.
- [6] Martens, E., Pestman, W., De Boer, A., Belitser, S., Klungel, O. (2006). *Instrumental Variables – Application and Limitations*. *Epidemiology*, 17(3), 260–267.
- [7] Höfler, M. (2005). *Causal Inference Based on Counterfactuals*. *BMC Medical Research Methodology*, 5(28).
- [8] Greenland, S., Robins, J., Pearl, J. (1999). *Confounding and Collapsibility in Causal Inference*. *Statistical Science*, 14(1), 29–46.
- [9] Hernán, M., Robins, J. (2006). *Instruments for Causal Inference – An Epidemiologist’s Dream?* *Epidemiology*, 17(4), 360–372.
- [10] Greenland, S. (2000). *An Introduction to Instrumental Variables for Epidemiologists*. *International Journal of Epidemiology*, 29, 722–729.
- [11] Relton, C., Davey Smith, G. (2012). *Two-step Epigenetic Mendelian Randomization: a Strategy for Establishing the Causal Role of Epigenetic Processes in Pathways to Disease*. *International Journal of Epidemiology*, 41, 161–176.
- [12] Zeng, Y., Cao, S., Yang, H. (2023). *The Causal Role of Gastroesophageal Reflux Disease in Anxiety Disorders and Depression: A Bidirectional Mendelian Randomization Study*. *Frontiers in Psychiatry*, 14.
- [13] Dudbridge, F. (2021). *Polygenic Mendelian Randomization*. *Cold Spring Harbor Perspectives in Medicine*, 11.

- [14] Lee, Y.H. (2020). *Overview of Mendelian Randomization Analysis*. Journal of Rheumatic Diseases Vol. 27, No. 4, 241–246.
- [15] Yavorska, O., Burgess, S. (2017). *MendelianRandomization: an R package for Performing Mendelian Randomization Analyses Using Summarized Data*. International Journal of Epidemiology, 46(6), 1734–1739.
- [16] Rovio, S., Pahkala, K., Nevalainen, J., Juonala, M., Salo, P., Kähönen, M., Hutri-Kähönen, N., Lehtimäki, T., Jokinen, E., Laitinen, T., Taittonen, L., Tossavainen, P., Viikari, J., Rinne, J., Raitakari, O. (2016). *Cognitive Performance in Young Adulthood and Midlife: Relations With Age, Sex, and Education—The Cardiovascular Risk in Young Finns Study*. Neuropsychology, 30(5), 532–542.
- [17] Ge, T., Chen, CY., Ni, Y., Feng, YC., Smoller, J. (2019). *Polygenic Prediction via Bayesian Regression and Continuous Shrinkage Priors*. Nature Communications, 10(1776).
- [18] Fox, J., Nie, Z., Byrnes, J., Culbertson, M., DebRoy, S., Friendly, M., Goodrich, B., Jones, R., Kramer, A., Monette, G., Novomestky, F. (2024). *sem: Structural Equation Models*. R-package version 3.1-16.
- [19] Burgess, S., Yavorska, O., Staley, J., Hartwig, F., Broadbent, J., Foley, C., Grant, A., Mason, A., Ye, T., Xue, H., Lin, Z., Xu, S., Patel, A., Kang, H., Wang, S., Karhunen, V. (2024). *MendelianRandomization: Mendelian Randomization Package*. R-package version 0.10.0.
- [20] Hakala, J., Pahkala, K., Juonala, M., Salo, P., Kähönen, M., Hutri-Kähönen, N., Lehtimäki, T., Laitinen, T., Jokinen, E., Taittonen, L., Tossavainen, P., Viikari, J., Raitakari, O., Rovio, S. (2021). *Cardiovascular Risk Factor Trajectories Since Childhood and Cognitive Performance in Midlife*. Circulation, 143, 1949–1961.

## A Liite: SAS-koodi

```
libname GRADU "/home/u63822714/gradu";

/*cognitive scores*/

data cantab;
set gradu.stualab11;
run;

proc sql;
create table cognitive as
select *
from cantab
where cantab11 eq 1;
quit;

data cognitivescores;
set cognitive;
keep Tutkno11 SP IKA80 COGNISCORE11 PALSCORE11 RTISCORE11
RVPSCORE11 SWMSCORE11 COGFAC11 PALFAC11 RTIFAC11 RVPFAC11
SWMFAC11;
run;

data gradu.cogniscores;
set cognitivescores;
run;

/*auc*/

data auc;
set gradu.aucauc;
run;

data auc1;
set auc;
keep tutkno01 syst_6_24_auc dkv_6_24_auc nonhdl_6_24_auc
ldl_6_24_auc bmi_6_24_auc;
run;

data gradu.auc;
set auc1;
run;
```

```

/*merge*/

data cognitivescores1;
set cognitivescores (rename=(tutkno11 = tutkno));
run;

data auc2;
set auc1 (rename=(tutkno01 = tutkno));
run;

data analysis1;
merge cognitivescores1 (in=a) auc2;
by tutkno;
if a;
run;

data gradu.analysisdata1;
set analysis1;
run;

data prs;
set gradu.yfs_g1_topmed_prscs_scores;
run;

data prscs;
set prs;
keep TUTKNO80 nonHDL_GLGC_prscs nonHDL_GLGC_prscs_SD
LDL_GLGC_prscs LDL_GLGC_prscs_SD SBP_UKBB_prscs
SBP_UKBB_prscs_SD DBP_UKBB_prscs DBP_UKBB_prscs_SD
BMI_UKBB_prscs BMI_UKBB_prscs_SD;
run;

data prscs;
set prscs (rename=(tutkno80 = tutkno));
run;

data gradu.prscs;
set prscs;
run;

proc sort data = analysis1;
by tutkno;
run;

```

```
data analysis_full;  
merge analysis1 (in=a) prscs;  
by tutkno;  
if a;  
run;
```

```
data gradu.analysisfull;  
set analysis_full;  
run;
```

## B Liite: R-koodi ja -tuloste

# Mendelistinen satunnaistaminen: kardiovaskulaaristen riskitekijöiden vaikutus kognitiiviseen toimintaan

Eetu Vento

Alustetaan tarvittavat kirjastot.

```
library(haven)
library(MendelianRandomization)
library(sem)
library(dplyr)
```

Ajetaan aineisto sisään ja tarkastellaan muuttujia.

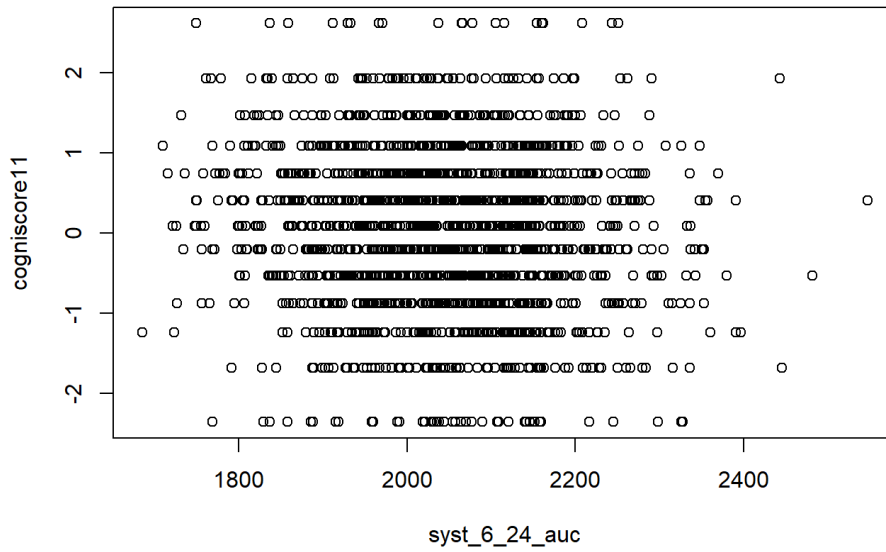
```
analysisfull <- read_sas("C:/Users/Käyttäjä/OneDrive - O365 Turun yliopisto/GRADU/AINEISTO/SAS/analysisfull.sas7bdat")
summary(analysisfull)
```

```
##      tutkno          SP          IKA80      cogniscore11
## Min.   : 1   Min.   :1.000   Min.   : 3.00   Min.   : -2.35621
## 1st Qu.:2195 1st Qu.:1.000   1st Qu.: 6.00   1st Qu.: -0.53187
## Median :3480 Median :1.000   Median :12.00   Median : 0.09175
## Mean   :3531 Mean   :1.455   Mean   :10.84   Mean   : 0.00000
## 3rd Qu.:5011 3rd Qu.:2.000   3rd Qu.:15.00   3rd Qu.: 0.74370
## Max.   :7073 Max.   :2.000   Max.   :18.00   Max.   : 2.61577
##                                     NA's   :228
##      palscore11      rtiscore11      rvpscore11      swmscore11
## Min.   : -3.39900   Min.   : -3.16646   Min.   : -2.20376   Min.   : -3.42205
## 1st Qu.: -0.63941   1st Qu.: -0.58216   1st Qu.: -0.64307   1st Qu.: -0.69244
## Median : 0.00475   Median : -0.08701   Median : -0.04826   Median : -0.03679
## Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.00000
## 3rd Qu.: 0.68351   3rd Qu.: 0.69709   3rd Qu.: 0.71305   3rd Qu.: 0.66435
## Max.   : 1.91317   Max.   : 2.30382   Max.   : 3.01128   Max.   : 2.12335
## NA's   :178       NA's   :204       NA's   :51       NA's   :15
##      cogfac11      palfac11      rtifac11      rvpfac11
## Min.   : -3.3915   Min.   : -3.399   Min.   : -3.395   Min.   : -3.417
## 1st Qu.: -0.6739   1st Qu.: -0.674   1st Qu.: -0.674   1st Qu.: -0.674
## Median : 0.0000   Median : 0.000   Median : 0.000   Median : 0.000
## Mean   : 0.0000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.000
## 3rd Qu.: 0.6739   3rd Qu.: 0.674   3rd Qu.: 0.674   3rd Qu.: 0.674
## Max.   : 3.3915   Max.   : 2.875   Max.   : 3.395   Max.   : 3.417
## NA's   :228       NA's   :178       NA's   :204       NA's   :51
##      swmfac11      bmi_6_24_auc      syst_6_24_auc      nonhdl_6_24_auc
## Min.   : -3.422   Min.   :254.8   Min.   :1685   Min.   : 14.03
## 1st Qu.: -0.674   1st Qu.:328.9   1st Qu.:1965   1st Qu.: 56.98
## Median : 0.000   Median :349.7   Median :2045   Median : 65.13
## Mean   : 0.000   Mean   :355.4   Mean   :2046   Mean   : 65.11
## 3rd Qu.: 0.674   3rd Qu.:375.1   3rd Qu.:2127   3rd Qu.: 71.89
## Max.   : 3.422   Max.   :647.3   Max.   :2547   Max.   :110.82
## NA's   :15
##      ldl_6_24_auc      dkv_6_24_auc      nonHDL_GLGC_prscs      LDL_GLGC_prscs
## Min.   :13.33   Min.   : 852.1   Min.   :0       Min.   :0
## 1st Qu.:50.22   1st Qu.:1135.0   1st Qu.:0       1st Qu.:0
## Median :57.92   Median :1193.4   Median :0       Median :0
## Mean   :58.21   Mean   :1184.5   Mean   :0       Mean   :0
## 3rd Qu.:65.32   3rd Qu.:1236.4   3rd Qu.:0       3rd Qu.:0
## Max.   :98.77   Max.   :1447.1   Max.   :0       Max.   :0
##                                     NA's   :211       NA's   :211
##      SBP_UKBB_prscs      DBP_UKBB_prscs      BMI_UKBB_prscs      nonHDL_GLGC_prscs_SD
## Min.   :0       Min.   :0       Min.   : -3.08524   Min.   : -4.04771
## 1st Qu.:0       1st Qu.:0       1st Qu.: -0.67648   1st Qu.: -0.67995
## Median :0       Median :0       Median : -0.00678   Median : 0.02510
## Mean   :0       Mean   :0       Mean   : 0.00591   Mean   : -0.01588
## 3rd Qu.:0       3rd Qu.:0       3rd Qu.: 0.69728   3rd Qu.: 0.70111
## Max.   :0       Max.   :0       Max.   : 3.21757   Max.   : 3.19891
## NA's   :211       NA's   :211       NA's   :211       NA's   :211
##      LDL_GLGC_prscs_SD      SBP_UKBB_prscs_SD      DBP_UKBB_prscs_SD      BMI_UKBB_prscs_SD
## Min.   : -3.87850   Min.   : -3.01831   Min.   : -3.37315   Min.   : -3.08362
## 1st Qu.: -0.68072   1st Qu.: -0.68836   1st Qu.: -0.67439   1st Qu.: -0.67526
## Median : 0.04074   Median : -0.03935   Median : 0.00939   Median : -0.00567
## Mean   : -0.01597   Mean   : -0.03840   Mean   : -0.01951   Mean   : 0.00702
## 3rd Qu.: 0.66677   3rd Qu.: 0.65710   3rd Qu.: 0.66320   3rd Qu.: 0.69828
## Max.   : 2.87171   Max.   : 3.70905   Max.   : 3.09588   Max.   : 3.21816
## NA's   :211       NA's   :211       NA's   :211       NA's   :211
```

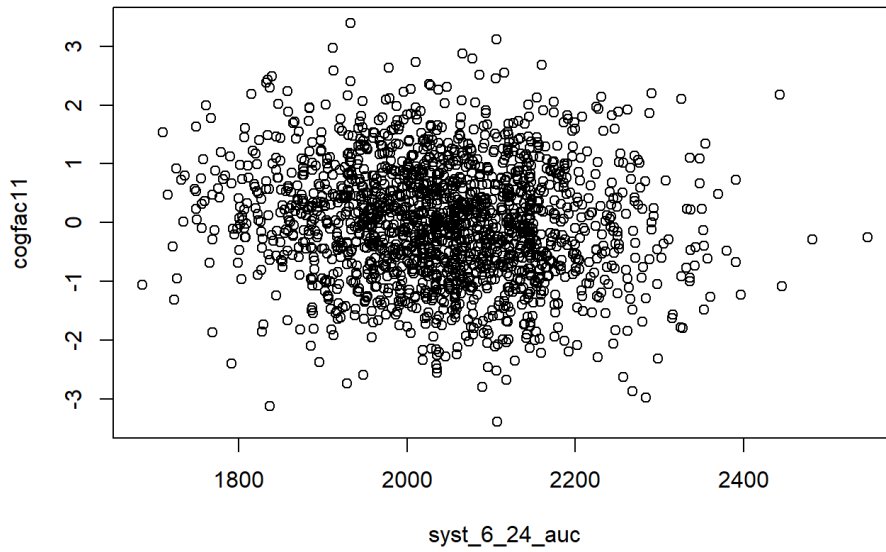
Vertaillaan vastemuuttujien arvoja score- ja fac-muuttujien osalta.

Kognitiivinen kokonaistesti ja systolinen verenpaine:

```
attach(analysisfull)
plot(syst_6_24_auc, cogniscore11)
```

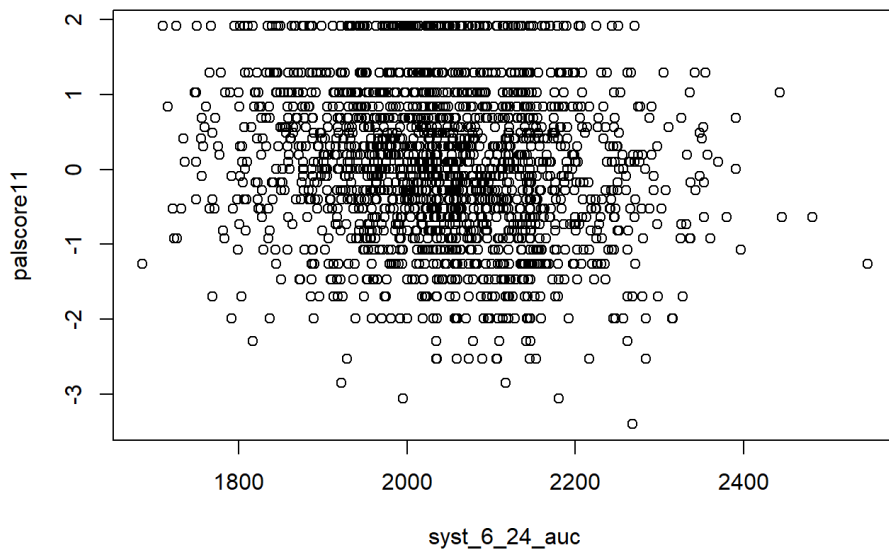


```
plot(syst_6_24_auc, cogfac11)
```

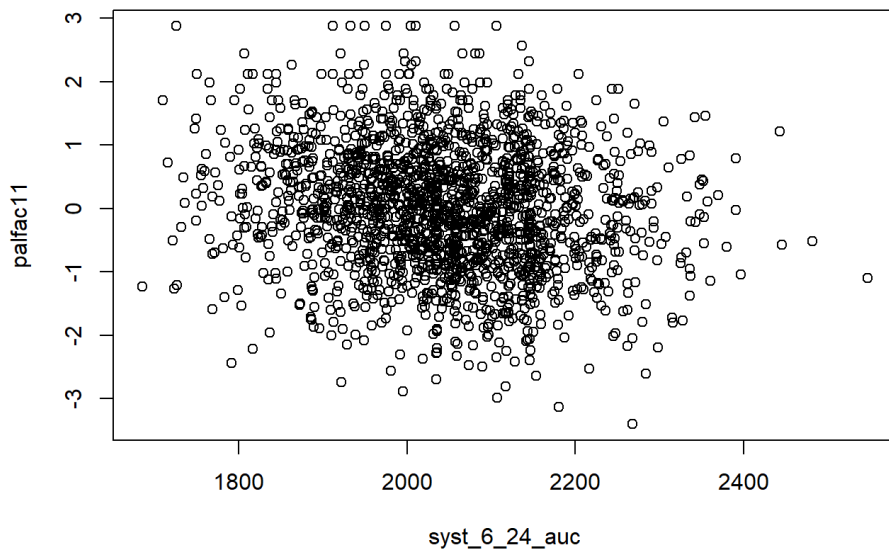


PAL-testi ja systolinen verenpaine:

```
plot(syst_6_24_auc, palscore11)
```

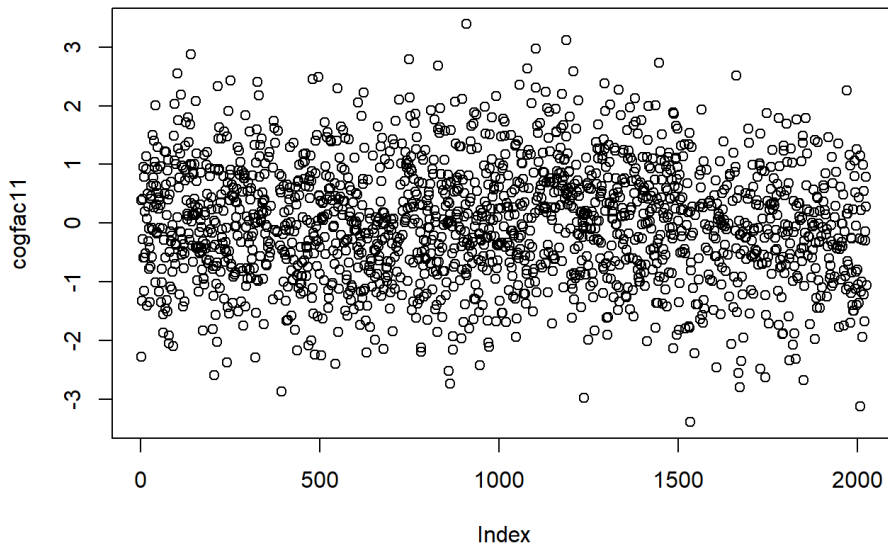


```
plot(syst_6_24_auc, palfac11)
```

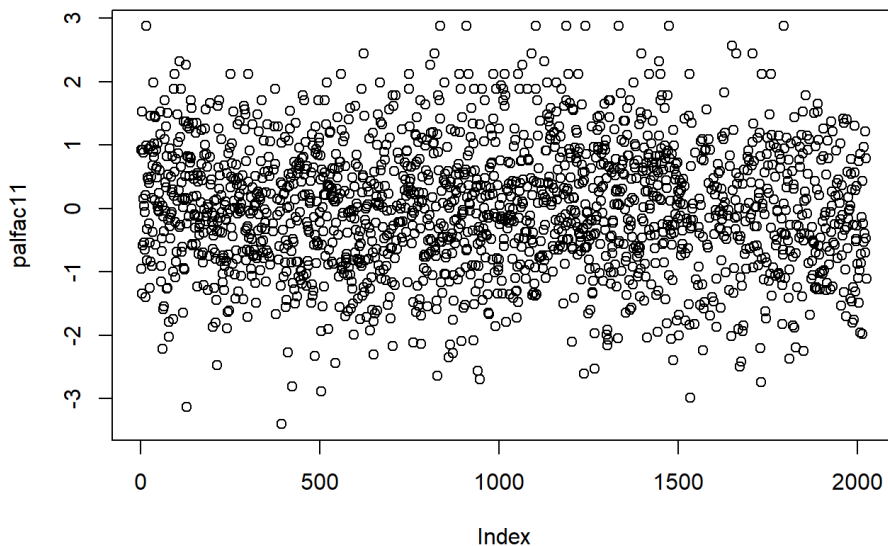


Pelkät fac-vastemuuttajat:

```
plot(cogfac11)
```



```
plot(palfac11)
```



Fac-vastemuuttuja eli pääkomponenttianalysillä muodostettu vastemuuttuja näyttää paremmalta, koska sen arvot näyttävät enemmän jatkuvilta kuin score-vastemuuttujan, jonka arvot näyttävät enemmän järjestysasteikollisilta. Fac-muuttujan arvoissa näyttäisi olevan siten myös enemmän variaatiota. Aiemman tutkimuksen tutkijat olivat myös tehneet saman huomion aikaisemmin ja siksi myös käyttäneet pääkomponenttianalysillä muodostettuja fac-vastemuuttuja analyseissä.

Tarkastellaan käytettävien muuttujien pituuksia. Tutkittavia on aineistossa 2026.

```
#PRSCS
sum(!is.na(SBP_UKBB_prscs_SD)) #1815
```

```
## [1] 1815
```

```
sum(!is.na(DBP_UKBB_prscs_SD)) #1815
```

```
## [1] 1815
```

```
sum(!is.na(nonHDL_GLGC_prscs_SD)) #1815
```

```
## [1] 1815
```

```
sum(!is.na(LDL_GLGC_prscs_SD)) #1815
```

```
## [1] 1815
```

```
sum(!is.na(BMI_UKBB_prscs_SD)) #1815
```

```
## [1] 1815
```

PRSCS-muuttujien eli instrumenttimuuttujien arvot löytyvät 1815 tutkittavalta.

```
#Exposures
sum(!is.na(syst_6_24_auc)) #2026
```

```
## [1] 2026
```

```
sum(!is.na(dkv_6_24_auc)) #2026
```

```
## [1] 2026
```

```
sum(!is.na(nonhdl_6_24_auc)) #2026
```

```
## [1] 2026
```

```
sum(!is.na(ldl_6_24_auc)) #2026
```

```
## [1] 2026
```

```
sum(!is.na(bmi_6_24_auc)) #2026
```

```
## [1] 2026
```

Altistemuuttujien arvot löytyvät kaikilta 2026 tutkittavalta.

```
#Outcomes
sum(!is.na(cogfac11)) #1798
```

```
## [1] 1798
```

```
sum(!is.na(palfac11)) #1848
```

```
## [1] 1848
```

```
sum(!is.na(rtifac11)) #1822
```

```
## [1] 1822
```

```
sum(!is.na(swmfac11)) #2011
```

```
## [1] 2011
```

```
sum(!is.na(rvpfac11)) #1975
```

```
## [1] 1975
```

```
detach(analysisfull)
```

Vastemuuttujien eli kognitiivisten testien arvot löytyvät 1798-2011 tutkittavalta testistä riippuen. Tämä luku vaihtelee, koska tutkittavat ovat saattaneet keskeyttää suorittamatta kaikkia testejä tai on ilmennyt teknisiä ongelmia. Jokaisen testin on tehnyt 1798 tutkittavaa.

Tarkastellaan kuvien avulla polygeenisten riskisummien yhteyttä altisteeseen, jotta nähdään, että niiden välillä on korrelaatiota sekä F-testisuure on vähintään 10.

```
#Systolic Blood Pressure
zx_syst <- lm(analysisfull$syst_6_24_auc~analysisfull$SBP_UKBB_prscs_SD)
summary(zx_syst)
```

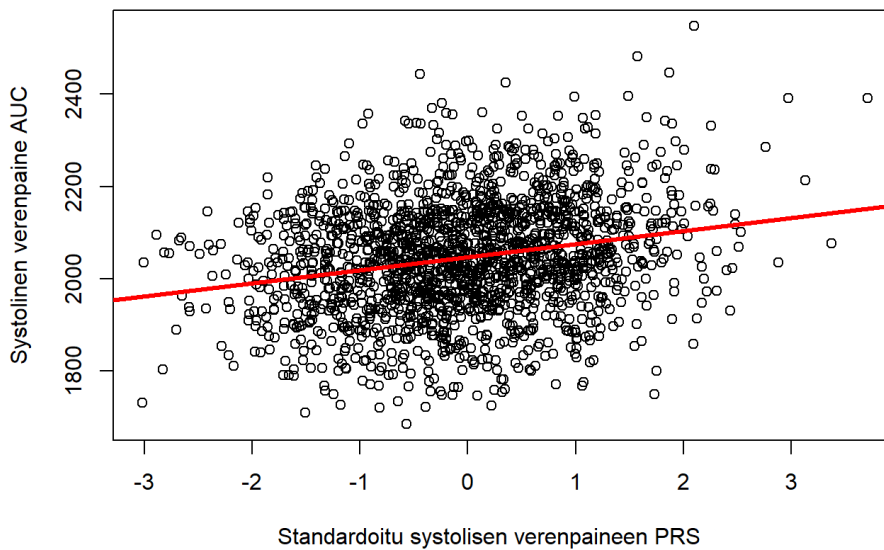
```
##
## Call:
## lm(formula = analysisfull$syst_6_24_auc ~ analysisfull$SBP_UKBB_prscs_SD)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -346.94  -77.22   -0.01   81.27  440.83
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2046.979      2.808  729.02 <2e-16 ***
## analysisfull$SBP_UKBB_prscs_SD  28.361      2.829   10.03 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 119.5 on 1813 degrees of freedom
## (211 observations deleted due to missingness)
## Multiple R-squared:  0.05254, Adjusted R-squared:  0.05202
## F-statistic: 100.5 on 1 and 1813 DF, p-value: < 2.2e-16
```

```
cor(analysisfull$syst_6_24_auc, analysisfull$SBP_UKBB_prscs_SD, use = "complete.obs")
```

```
## [1] 0.2292172
```

```
plot(zx_syst$model$`analysisfull$SBP_UKBB_prscs_SD`, zx_syst$model$`analysisfull$syst_6_24_auc`, main="Polygeenisen riskisumman yhteys altisteeseen", xlab="Standardoitu systolisen verenpaineen PRSCS", ylab="Systolinen verenpaine AUC")
abline(zx_syst, col="red", lwd=3)
```

### Polygeenisen riskisumman yhteys altisteeseen



```
#Diastolic Blood Pressure
```

```
zx_dias <- lm(analysisfull$dkv_6_24_auc~analysisfull$DBP_UKBB_prscs_SD)
summary(zx_dias)
```

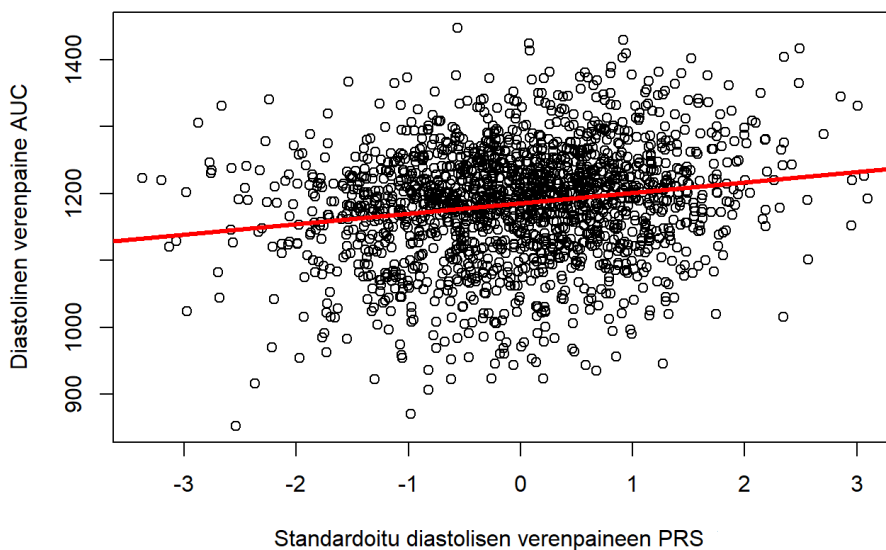
```
##
## Call:
## lm(formula = analysisfull$dkv_6_24_auc ~ analysisfull$DBP_UKBB_prscs_SD)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -299.73  -51.69   8.18   55.06  270.79
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1185.122     1.966 602.942 < 2e-16 ***
## analysisfull$DBP_UKBB_prscs_SD  15.635     1.979   7.899 4.81e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 83.72 on 1813 degrees of freedom
## (211 observations deleted due to missingness)
## Multiple R-squared:  0.03327, Adjusted R-squared:  0.03274
## F-statistic: 62.4 on 1 and 1813 DF, p-value: 4.815e-15
```

```
cor(analysisfull$dkv_6_24_auc, analysisfull$DBP_UKBB_prscs_SD, use = "complete.obs")
```

```
## [1] 0.1824109
```

```
plot(zx_dias$model$`analysisfull$DBP_UKBB_prscs_SD`, zx_dias$model$`analysisfull$dkv_6_24_auc`, main="Polygeenisen riskisumman yhteys altisteeseen", xlab="Standardoitu diastolisen verenpaineen PRSCS", ylab="Diastolinen verenpaine AUC" )
abline(zx_dias, col="red", lwd=3)
```

### Polygeenisen riskisumman yhteys altisteeseen



```
#NonHDL
```

```
zx_nonhdl <- lm(analysisfull$nonhdl_6_24_auc~analysisfull$nonHDL_GLGC_prscs_SD)
summary(zx_nonhdl)
```

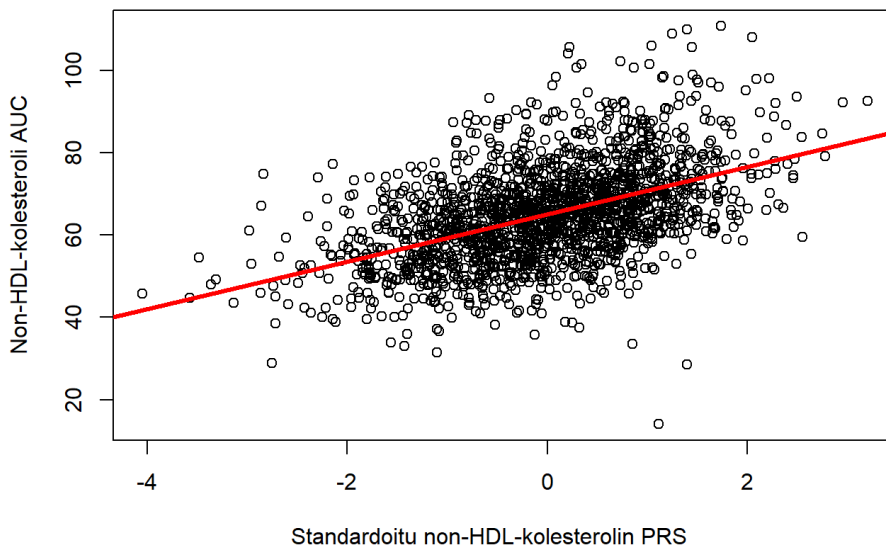
```
##
## Call:
## lm(formula = analysisfull$nonhdl_6_24_auc ~ analysisfull$nonHDL_GLGC_prscs_SD)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -57.530  -6.867  -0.522   5.739  39.230
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      65.1467     0.2445  266.44 <2e-16 ***
## analysisfull$nonHDL_GLGC_prscs_SD  5.7672     0.2441   23.63 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.42 on 1813 degrees of freedom
## (211 observations deleted due to missingness)
## Multiple R-squared:  0.2355, Adjusted R-squared:  0.2351
## F-statistic: 558.4 on 1 and 1813 DF,  p-value: < 2.2e-16
```

```
cor(analysisfull$nonhdl_6_24_auc, analysisfull$nonHDL_GLGC_prscs_SD, use = "complete.obs")
```

```
## [1] 0.485259
```

```
plot(zx_nonhdl$model$`analysisfull$nonHDL_GLGC_prscs_SD`, zx_nonhdl$model$`analysisfull$nonhdl_6_24_auc`, main = "Polygeenis
en riskisumman yhteys altisteeseen", xlab="Standardoitu non-HDL-kolesterolin PRSCS", ylab="Non-HDL-kolesteroli AUC")
abline(zx_nonhdl, col="red", lwd=3)
```

### Polygeenis en riskisumman yhteys altisteeseen



```
#LDL
```

```
zx_LDL <- lm(analysisfull$ldl_6_24_auc~analysisfull$LDL_GLGC_prscs_SD)
summary(zx_LDL)
```

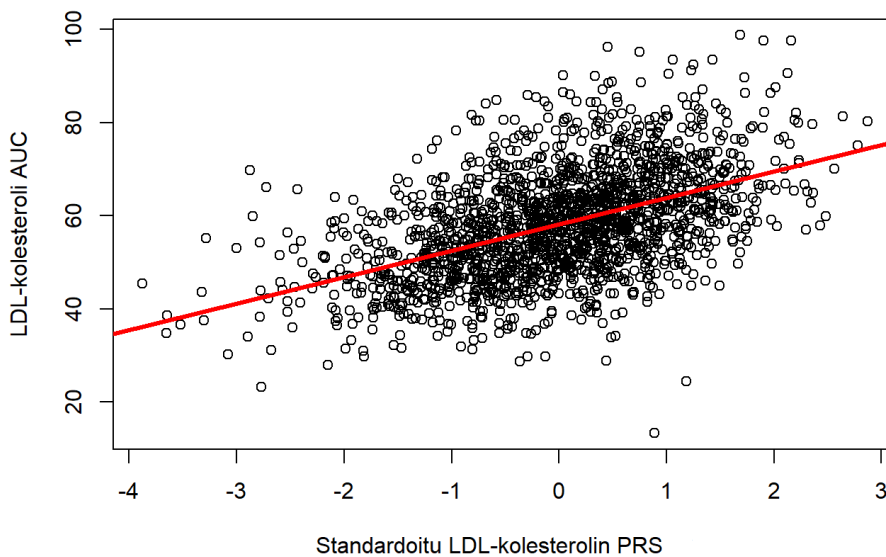
```
##
## Call:
## lm(formula = analysisfull$ldl_6_24_auc ~ analysisfull$LDL_GLGC_prscs_SD)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -49.869  -6.777  -0.656   5.936  35.498
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      58.1747     0.2363   246.1 <2e-16 ***
## analysisfull$LDL_GLGC_prscs_SD  5.6726     0.2354    24.1 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.07 on 1813 degrees of freedom
## (211 observations deleted due to missingness)
## Multiple R-squared:  0.2426, Adjusted R-squared:  0.2422
## F-statistic: 580.8 on 1 and 1813 DF,  p-value: < 2.2e-16
```

```
cor(analysisfull$ldl_6_24_auc, analysisfull$LDL_GLGC_prscs_SD, use = "complete.obs")
```

```
## [1] 0.4925693
```

```
plot(zx_LDL$model$`analysisfull$LDL_GLGC_prscs_SD`, zx_LDL$model$`analysisfull$ldl_6_24_auc`, main = "Polygeenisen riskisumman yhteys altisteeseen", xlab="Standardoitu LDL-kolesterolin PRSCS", ylab="LDL-kolesteroli AUC")
abline(zx_LDL, col="red", lwd=3)
```

### Polygeenisen riskisumman yhteys altisteeseen



```
#BMI
```

```
zx_BMI <- lm(analysisfull$bmi_6_24_auc~analysisfull$BMI_UKBB_prscs_SD)
summary(zx_BMI)
```

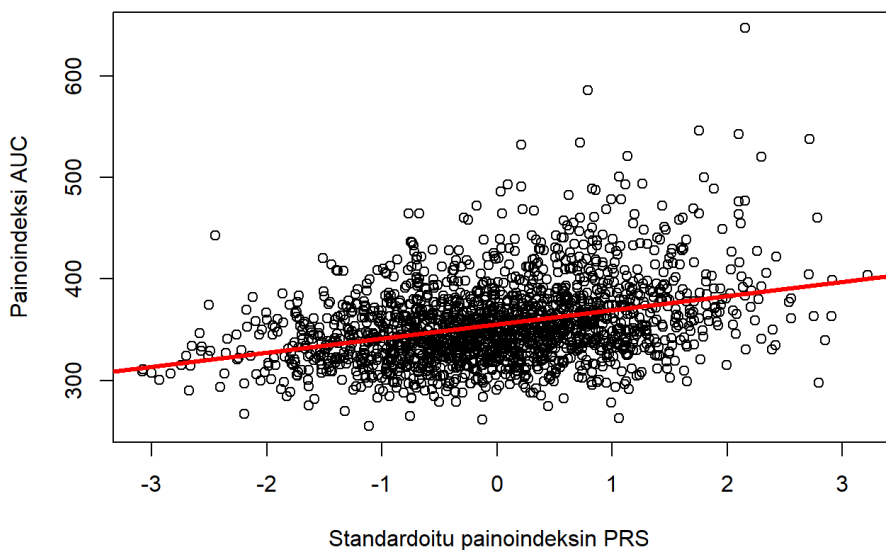
```
##
## Call:
## lm(formula = analysisfull$bmi_6_24_auc ~ analysisfull$BMI_UKBB_prscs_SD)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -107.448  -24.538   -4.772   18.495  262.050
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      355.3651     0.8893   399.61 <2e-16 ***
## analysisfull$BMI_UKBB_prscs_SD  13.8973     0.8941   15.54 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.89 on 1813 degrees of freedom
## (211 observations deleted due to missingness)
## Multiple R-squared:  0.1176, Adjusted R-squared:  0.1171
## F-statistic: 241.6 on 1 and 1813 DF,  p-value: < 2.2e-16
```

```
cor(analysisfull$bmi_6_24_auc, analysisfull$BMI_UKBB_prscs_SD, use = "complete.obs")
```

```
## [1] 0.3429193
```

```
plot(zx_BMI$model$`analysisfull$BMI_UKBB_prscs_SD`, zx_BMI$model$`analysisfull$bmi_6_24_auc`, main = "Polygeenisen riskisumman yhteys altisteseen", xlab="Standardoitu painoindeksi PRSCS", ylab="Painoindeksi AUC")
abline(zx_BMI, col="red", lwd=3)
```

### Polygeenisen riskisumman yhteys altisteseen



Instrumentti- ja altistemuuttujien suhteet vaikuttavat korreloituneilta eli niiden välillä on yhteys, mikä on yksi instrumenttimuuttujan täytettävistä vaatimuksista. Verenpaineiden osalta korrelaatio on heikointa (noin 0.2) ja kolesterolin osalta vahvinta (noin 0.5). Myös F-testisuureen arvot ovat suuria ja helposti yli 10.

## Systolinen verenpaine

Luodaan aluksi systoliselle verenpaineelle tulosmatriisi, johon osatestin piste-estimaatit kootaan vertailun vuoksi.

```
#Create a datamatrix for results

# create matrix with 1 column and 5 rows
RESULTS_SystolicBP= as.data.frame(matrix(rep(0,5), ncol=1, byrow=TRUE))

# specify the column names and row names of matrix
colnames(RESULTS_SystolicBP) = c('Effect')
rownames(RESULTS_SystolicBP) <- c('Basic LM', 'MR 2SLS1', 'MR 2SLS2', 'MR IVW', 'MR TRIANGLE')
RESULTS_SystolicBP
```

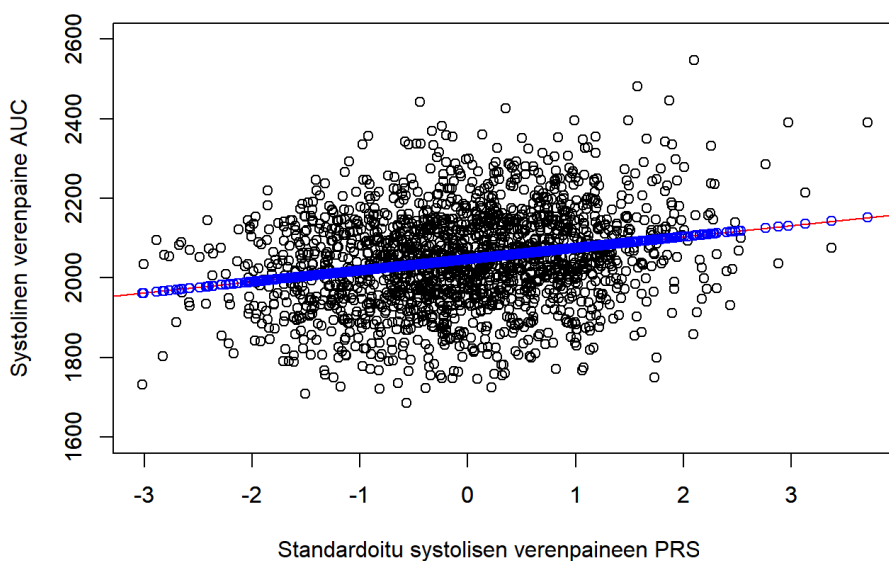
```
##          Effect
## Basic LM      0
## MR 2SLS1      0
## MR 2SLS2      0
## MR IVW        0
## MR TRIANGLE   0
```

## PAL-testi

Kaksivaiheinen pienimmän neliösumman menetelmä askel askeleelta:

```
#2SLS step by step
syst <- glm(syst_6_24_auc~SBP_UKBB_prscs_SD, data = analysisfull)
```

```
plot(syst$model$SBP_UKBB_prscs_SD, syst$model$syst_6_24_auc, ylim = c(1600,2600), xlab="Standardoitu systolisen verenpaineen PRSCS", ylab="Systolinen verenpaine AUC")
abline(syst, col="red")
par(new=T)
plot(syst$model$SBP_UKBB_prscs_SD, syst$fitted.values, col="blue", ylim = c(1600,2600), ylab = "", xlab = "")
```



Polygeenisen riskisumman avulla ennustetut altisteen arvot näkyvät sinisellä.

```
syst_pred <- predict(syst, newdata = analysisfull)
analysisfull$fitted <- syst_pred
length(syst_pred) #same Length 1848 as data
```

```
## [1] 2026
```

```
m1 <- glm(palfac11~syst_pred, data = analysisfull) #predicted
m2 <- glm(palfac11~syst_6_24_auc, data = analysisfull) #original
summary(m1) #predicted
```

```
##
## Call:
## glm(formula = palfac11 ~ syst_pred, data = analysisfull)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3962  -0.6767  -0.0006   0.6701   2.8782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.028e-02  1.803e+00  -0.017   0.987
## syst_pred    1.339e-05  8.811e-04   0.015   0.988
##
## (Dispersion parameter for gaussian family taken to be 1.001922)
##
##      Null deviance: 1656.2  on 1654  degrees of freedom
## Residual deviance: 1656.2  on 1653  degrees of freedom
## (371 observations deleted due to missingness)
## AIC: 4703.9
##
## Number of Fisher Scoring iterations: 2
```

```
summary(m2) #original
```

```
##
## Call:
## glm(formula = palfac11 ~ syst_6_24_auc, data = analysisfull)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.11459  -0.65544  -0.00639   0.65933   2.95393
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.5947696  0.3840462   6.756 1.89e-11 ***
## syst_6_24_auc -0.0012697  0.0001876  -6.769 1.74e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9741017)
##
##      Null deviance: 1842.8  on 1847  degrees of freedom
## Residual deviance: 1798.2  on 1846  degrees of freedom
## (178 observations deleted due to missingness)
## AIC: 5199.9
##
## Number of Fisher Scoring iterations: 2
```

Ennustetun mallin 95 prosentin luottamusväli regressiokertoimen estimaatille:

```
confint(m1, level = 0.95)
```

```
##              2.5 %      97.5 %
## (Intercept) -3.563464146  3.502906730
## syst_pred    -0.001713602  0.001740388
```

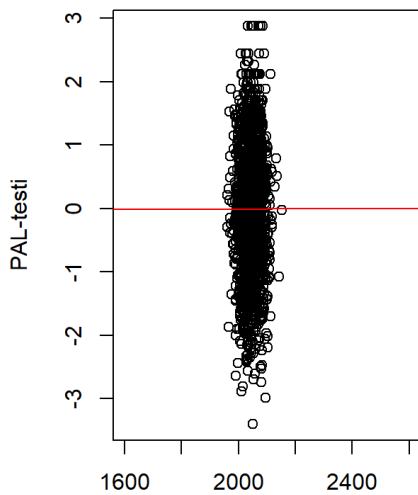
Lineaarisen eli "alkuperäisen" mallin 95 prosentin luottamusväli regressiokertoimen estimaatille:

```
confint(m2, level = 0.95)
```

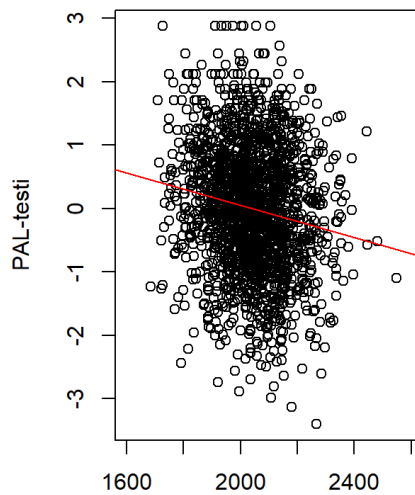
```
##              2.5 %      97.5 %
## (Intercept)  1.842052894  3.3474862883
## syst_6_24_auc -0.001637321 -0.0009020055
```

```
RESULTS_SystolicBP[2,1] <- m1$coefficients[2]
RESULTS_SystolicBP[1,1] <- m2$coefficients[2]
```

```
par(mfrow=c(1,2))
plot(m1$model$syst_pred, m1$model$palfac11, ylab = "PAL-testi", xlab = "Ennustettu systolinen verenpaine AUC", xlim=c(1600,2600)) #pred
abline(m1, col="red")
plot(m2$model$syst_6_24_auc, m2$model$palfac11, ylab = "PAL-testi", xlab = "Alkuperäinen systolinen verenpaine AUC", xlim = c(1600,2600)) #orig
abline(m2, col="red")
```



Ennustettu systolinen verenpaine AUC



Alkuperäinen systolinen verenpaine AUC

```
par(mfrow=c(1,1))
```

Verrataan R:n sisäänrakennettuun 2SLS-funktioon.

```
#compare to build-in-function
tsls_syst <- tsls(palfac11~syst_6_24_auc, ~SBP_UKBB_prscs_SD, data = analysisfull)
summary(tsls_syst) #same results
```

```
##
## 2SLS Estimates
##
## Model Formula: palfac11 ~ syst_6_24_auc
##
## Instruments: ~SBP_UKBB_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.399097 -0.677690  0.000429  0.000000  0.669838  2.882208
##
##              Estimate   Std. Error  t value Pr(>|t|)
## (Intercept) -2.999225e-02  1.784281e+00 -0.01681  0.98659
## syst_6_24_auc  1.326651e-05  8.730201e-04  0.01520  0.98788
##
## Residual standard error: 1.0012111 on 1653 degrees of freedom
```

```
confint(tsls_syst)
```

```
##              2.5 %      97.5 %
## (Intercept) -3.527118024  3.467133533
## syst_6_24_auc -0.001697821  0.001724354
```

```
RESULTS_SystolicBP[3,1] <- tsls_syst$coefficients[2]
```

Painotettu käänteisen varianssin menetelmä:

```
#inverse variance weighted method step by step

zx_syst <- glm(syst_6_24_auc~SBP_UKBB_prscs_SD, data = analysisfull)
zy_syst <- glm(palfac11~SBP_UKBB_prscs_SD, data = analysisfull)
s1 <- summary(zx_syst)
beta_syst_x <- s1$coefficients[2,1]
se_syst_x <- s1$coefficients[2,2]
s2 <- summary(zy_syst)
beta_syst_y <- s2$coefficients[2,1]
se_syst_y <- s2$coefficients[2,2]
mrdata_syst <- mr_input(bx = beta_syst_x, bxse = se_syst_x, by = beta_syst_y, byse = se_syst_y)

m <- mr_ivw(mrdata_syst)
m
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI p-value
## IVW 0.000 0.001 -0.002, 0.002 0.988
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 100.5.
```

```
RESULTS_SystolicBP[4,1] <- m$Estimate
#mr_plot(mrdata_syst)
```

Yksinkertainen kolmiomenetelmä/suhde-estimaattori:

```
# simple triangle method

RESULTS_SystolicBP[5,1] <- beta_syst_y/beta_syst_x
RESULTS_SystolicBP
```

```
##              Effect
## Basic LM    -1.269663e-03
## MR 2SLS1    1.339320e-05
## MR 2SLS2    1.326651e-05
## MR IVW      1.339320e-05
## MR TRIANGLE 1.339320e-05
```

Rakennetaan valmis funktio 2SLS-menetelmälle tulevia tarkasteluja varten.

```
#Building a two-stage least squares (2SLS) estimator function for advance

twoSLS <- function(xvar, xname, xpred, zvar, zname, yvar, yname, d){
  glmmodel <- glm(xvar~zvar)

  xvar_pred <- predict(glmmodel, newdata = d)
  d$xpred <- xvar_pred

  m1 <- glm(yvar~xpred, data = d)
  m2 <- glm(yvar~xvar, data = d)
  print(summary(m1))
  print(confint(m1))
  print(summary(m2))
  print(confint(m2))

  par(mfrow=c(1,2))
  plot(m1$model$xpred, m1$model$yvar, ylab= yname, xlab=xname, main="Altisteen kausaalinen vaikutus")
  abline(m1, col="red", lwd=3)
  plot(m2$model$xvar, m2$model$yvar, ylab=yname, xlab=xname, main="Alkuperäinen altisteen yhteys")
  abline(m2, col="red", lwd=3)
  par(mfrow=c(1,1))

  #Save results
  r = as.data.frame(matrix(rep(NA,10), ncol=2, byrow=TRUE))

  # specify the column names and row names of matrix
  colnames(r) = c('Test', 'Effect')
  r$Test <- c('Basic LM', 'MR 2SLS1', 'MR 2SLS2', 'MR IVW', 'MR TRIANGLE')

  r[1,2] <- m2$coefficients[2]
  r[2,2] <- m1$coefficients[2]

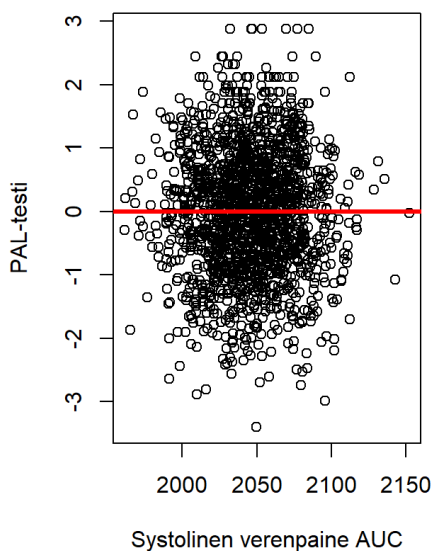
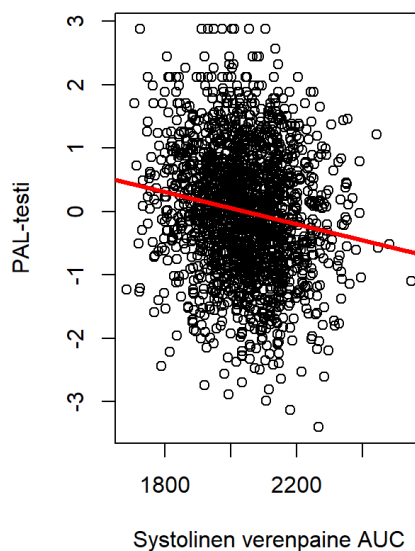
  return(r)
}
```

Verrataan aikaisempiin tuloksiin funktion validoimiseksi.

```
#Compare to previous results to validate function

r_syst1 <- twoSLS(xvar=analysisfull$syst_6_24_auc, xname = "Systolinen verenpaine AUC", xpred= "syst_pred",
  zvar=analysisfull$SBP_UKBB_prscs_SD, zname = "Standardoitu systolisen verenpaineen PRSCS",
  yvar=analysisfull$palfac11, yname="PAL-testi", d = analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3962  -0.6767  -0.0006   0.6701   2.8782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.028e-02  1.803e+00  -0.017   0.987
## xpred        1.339e-05  8.811e-04   0.015   0.988
##
## (Dispersion parameter for gaussian family taken to be 1.001922)
##
## Null deviance: 1656.2 on 1654 degrees of freedom
## Residual deviance: 1656.2 on 1653 degrees of freedom
## (371 observations deleted due to missingness)
## AIC: 4703.9
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -3.563464146  3.502906730
## xpred        -0.001713602  0.001740388
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.11459  -0.65544  -0.00639   0.65933   2.95393
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.5947696  0.3840462   6.756 1.89e-11 ***
## xvar        -0.0012697  0.0001876  -6.769 1.74e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9741017)
##
## Null deviance: 1842.8 on 1847 degrees of freedom
## Residual deviance: 1798.2 on 1846 degrees of freedom
## (178 observations deleted due to missingness)
## AIC: 5199.9
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  1.842052894  3.3474862883
## xvar        -0.001637321 -0.0009020055
```

**Altisteen kausaalinen vaikutus****Alkuperäinen altisteen yhteys**

Rakennetaan funktio painotetulle käänteisen varianssin menetelmälle jatkoa varten.

```
#Building an inverse variance weighted (IVW) method function for advance
```

```
inverse_variance_weighted <- function(xvar, zvar, yvar){
  zx <- glm(xvar~zvar)
  zy <- glm(yvar~zvar)
  s1 <- summary(zx)
  beta_x <- s1$coefficients[2,1]
  se_x <- s1$coefficients[2,2]
  s2 <- summary(zy)
  beta_y <- s2$coefficients[2,1]
  se_y <- s2$coefficients[2,2]
  mrdata <- mr_input(bx = beta_x, bxse = se_x, by = beta_y, byse = se_y)

  print(mr_ivw(mrdata))
  print(beta_y/beta_x)

  #Save results
  r = as.data.frame(matrix(rep(NA,10), ncol=2, byrow=TRUE))

  # specify the column names and row names of matrix
  colnames(r) = c('Test','Effect')
  r$Test <- c('Basic LM','MR 2SLS1','MR 2SLS2','MR IVW','MR TRIANGLE')

  r[4,2] <- mr_ivw(mrdata)$Estimate
  r[5,2] <- beta_y/beta_x
  return(r)
}
```

Verrataan aikaisemmin saatuihin tuloksiin funktion validoimiseksi.

```
#Compare to previous results to validate function
```

```
r_syst2 <- inverse_variance_weighted(xvar= analysisfull$syst_6_24_auc,
  zvar = analysisfull$SBP_UKBB_prscs_SD,
  yvar = analysisfull$palfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI p-value
## IVW 0.000 0.001 -0.002, 0.002 0.988
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 100.5.
## [1] 1.33932e-05
```

Yhdistetään tulokset:

```
#Combine Results
```

```
RESULTS_SystolicBP <- rows_patch(r_syst1, r_syst2, by="Test")
RESULTS_SystolicBP[3,2] <- tsls_syst$coefficients[2]
RESULTS_SystolicBP
```

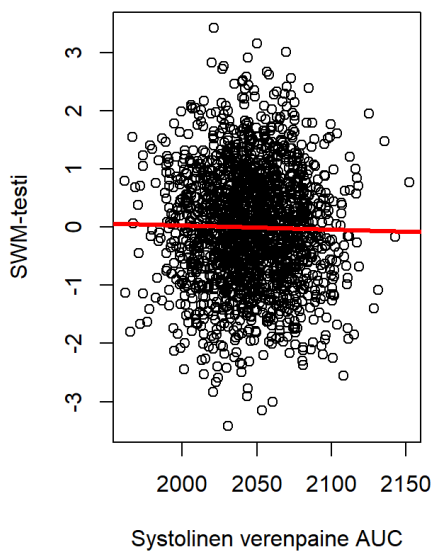
```
##          Test          Effect
## 1 Basic LM -1.269663e-03
## 2 MR 2SLS1 1.339320e-05
## 3 MR 2SLS2 1.326651e-05
## 4 MR IVW 1.339320e-05
## 5 MR TRIANGLE 1.339320e-05
```

## SWM-testi

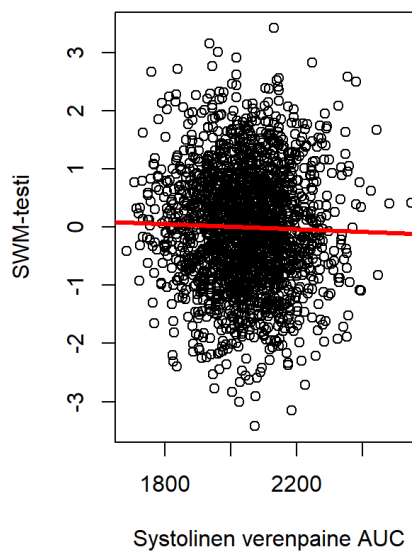
```
r_syst_swm1 <- twoSLS(xvar=analysisfull$syst_6_24_auc, xname = "Systolinen verenpaine AUC", xpred= "syst_pred",
  zvar=analysisfull$SBP_UKBB_prscs_SD, zname = "Standardoitu systolisen verenpaineen PRSCS",
  yvar=analysisfull$swmfac11, yname="SWM-testi", d = analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4328  -0.6698  -0.0060   0.6724   3.4047
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.3909928  1.7108766   0.813   0.416
## xpred        -0.0006796  0.0008361  -0.813   0.416
##
## (Dispersion parameter for gaussian family taken to be 0.9957586)
##
## Null deviance: 1792.0  on 1800  degrees of freedom
## Residual deviance: 1791.4  on 1799  degrees of freedom
## (225 observations deleted due to missingness)
## AIC: 5107.4
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -1.962263733  4.7442493045
## xpred        -0.002318397  0.0009592115
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4154  -0.6803  -0.0026   0.6776   3.4418
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.4702742  0.3740763   1.257   0.209
## xvar        -0.0002299  0.0001826  -1.259   0.208
##
## (Dispersion parameter for gaussian family taken to be 0.9979404)
##
## Null deviance: 2006.4  on 2010  degrees of freedom
## Residual deviance: 2004.9  on 2009  degrees of freedom
## (15 observations deleted due to missingness)
## AIC: 5706.8
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.2629018878  1.2034502630
## xvar        -0.0005877063  0.0001278913
```

Altisteen kausaalinen vaikutus



Alkuperäinen altisteen yhteys



```
tsls_syst_swm <- tsls(swmfac11~syst_6_24_auc, ~SBP_UKBB_prscs_SD, data = analysisfull)
summary(tsls_syst_swm)
```

```
##
## 2SLS Estimates
##
## Model Formula: swmfac11 ~ syst_6_24_auc
##
## Instruments: ~SBP_UKBB_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.403087 -0.673050 -0.006831  0.000000  0.656673  3.480046
##
##              Estimate   Std. Error  t value Pr(>|t|)
## (Intercept)  1.3982941798  1.7215039968  0.81225  0.41675
## syst_6_24_auc -0.0006832392  0.0008414303 -0.81200  0.41690
##
## Residual standard error: 0.9988315 on 1799 degrees of freedom
```

```
confint(tsls_syst_swm)
```

```
##              2.5 %      97.5 %
## (Intercept) -1.975791653  4.7723800128
## syst_6_24_auc -0.002332412  0.0009659339
```

```
r_syst_swm2 <- inverse_variance_weighted(xvar= analysisfull$syst_6_24_auc,
                                         zvar = analysisfull$SBP_UKBB_prscs_SD,
                                         yvar = analysisfull$swmfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
## IVW   -0.001    0.001 -0.002, 0.001  0.416
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 100.5.
## [1] -0.0006795929
```

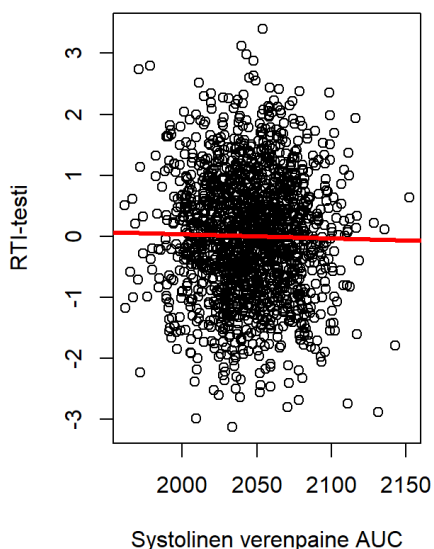
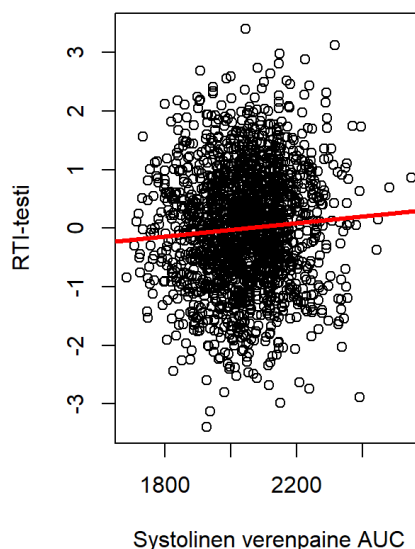
```
RESULTS_SystolicBP <- rows_patch(r_syst_swm1, r_syst_swm2, by="Test")
RESULTS_SystolicBP[3,2] <- tsls_syst_swm$coefficients[2]
RESULTS_SystolicBP
```

```
##      Test      Effect
## 1 Basic LM -0.0002299075
## 2 MR 2SLS1 -0.0006795929
## 3 MR 2SLS2 -0.0006832392
## 4 MR IVW -0.0006795929
## 5 MR TRIANGLE -0.0006795929
```

## RTI-testi

```
r_syst_rti1 <- twoSLS(xvar=analysisfull$syst_6_24_auc, xname = "Systolinen verenpaine AUC", xpred= "syst_pred",
                    zvar=analysisfull$SBP_UKBB_prscs_SD, zname = "Standardoitu systolisen verenpaineen PRSCS",
                    yvar=analysisfull$rtifac11, yname="RTI-testi", d = analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1392  -0.6818   0.0055   0.6715   3.3940
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.4215668  1.8127077   0.784   0.433
## xpred        -0.0006916  0.0008860  -0.781   0.435
##
## (Dispersion parameter for gaussian family taken to be 1.000386)
##
## Null deviance: 1629.2 on 1629 degrees of freedom
## Residual deviance: 1628.6 on 1628 degrees of freedom
## (396 observations deleted due to missingness)
## AIC: 4630.4
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -2.131275050  4.97440868
## xpred        -0.002428188  0.00104497
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3290  -0.6792   0.0031   0.6677   3.3946
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.1595740  0.3912264  -2.964  0.00308 **
## xvar         0.0005674  0.0001911   2.969  0.00302 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.993808)
##
## Null deviance: 1817.5 on 1821 degrees of freedom
## Residual deviance: 1808.7 on 1820 degrees of freedom
## (204 observations deleted due to missingness)
## AIC: 5163.3
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -1.9263637092 -0.3927842095
## xvar         0.0001928575  0.0009418962
```

**Altisteen kausaalinen vaikutus****Alkuperäinen altisteen yhteys**

```
tsls_syst_rti <- tsls(rtifac11~syst_6_24_auc, ~SBP_UKBB_prscs_SD, data = analysisfull)
summary(tsls_syst_rti)
```

```
##
## 2SLS Estimates
##
## Model Formula: rtifac11 ~ syst_6_24_auc
##
## Instruments: ~SBP_UKBB_prscs_SD
##
## Residuals:
##      Min.    1st Qu.    Median      Mean   3rd Qu.    Max.
## -3.203173 -0.683358 -0.005216  0.000000  0.682092  3.388971
##
##              Estimate   Std. Error  t value Pr(>|t|)
## (Intercept)  1.4188111172  1.8242429481  0.77775  0.43683
## syst_6_24_auc -0.0006909213  0.0008925155 -0.77413  0.43897
##
## Residual standard error: 1.0085216 on 1628 degrees of freedom
```

```
confint(tsls_syst_rti)
```

```
##              2.5 %      97.5 %
## (Intercept) -2.156639360  4.994261595
## syst_6_24_auc -0.002440219  0.001058377
```

```
r_syst_rti2 <- inverse_variance_weighted(xvar= analysisfull$syst_6_24_auc,
                                       zvar = analysisfull$SBP_UKBB_prscs_SD,
                                       yvar = analysisfull$rtifac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error  95% CI      p-value
##   IVW   -0.001    0.001 -0.002, 0.001  0.435
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 100.5.
## [1] -0.000691609
```

```
RESULTS_SystolicBP <- rows_patch(r_syst_rti1, r_syst_rti2, by="Test")
RESULTS_SystolicBP[3,2] <- tsls_syst_rti$coefficients[2]
RESULTS_SystolicBP
```

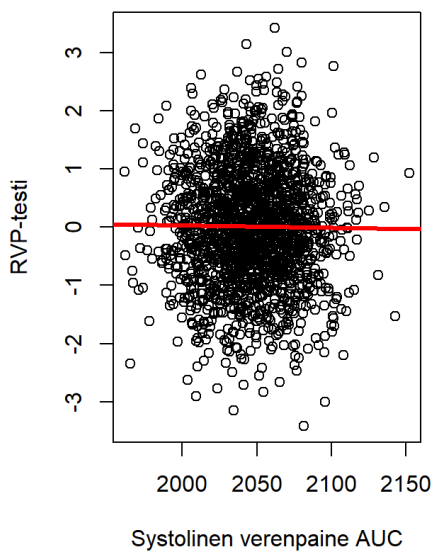
```
##      Test      Effect
## 1 Basic LM  0.0005673769
## 2 MR 2SLS1 -0.0006916090
## 3 MR 2SLS2 -0.0006909213
## 4 MR IVW   -0.0006916090
## 5 MR TRIANGLE -0.0006916090
```

## RVP-testi

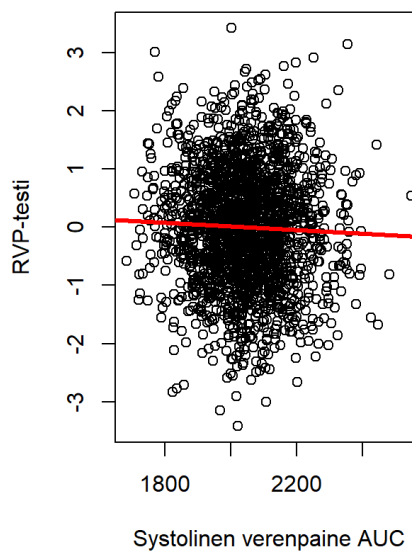
```
r_syst_rvp1 <- twoSLS(xvar=analysisfull$syst_6_24_auc, xname = "Systolinen verenpaine AUC", xpred= "syst_pred",
                    zvar=analysisfull$SBP_UKBB_prscs_SD, zname = "Standardoitu systolisen verenpaineen PRSCS",
                    yvar=analysisfull$rvpfac11, yname="RVP-testi", d = analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4152  -0.6688  -0.0038   0.6701   3.4108
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.8736123  1.7255248   0.506   0.613
## xpred       -0.0004206  0.0008433  -0.499   0.618
##
## (Dispersion parameter for gaussian family taken to be 0.997963)
##
## Null deviance: 1763.6 on 1768 degrees of freedom
## Residual deviance: 1763.4 on 1767 degrees of freedom
## (257 observations deleted due to missingness)
## AIC: 5020.6
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -2.508354105  4.255578689
## xpred       -0.002073436  0.001232195
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4245  -0.6691  -0.0016   0.6724   3.4039
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.6174386  0.3783614   1.632   0.103
## xvar        -0.0003018  0.0001846  -1.635   0.102
##
## (Dispersion parameter for gaussian family taken to be 0.9973585)
##
## Null deviance: 1970.5 on 1974 degrees of freedom
## Residual deviance: 1967.8 on 1973 degrees of freedom
## (51 observations deleted due to missingness)
## AIC: 5603.6
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.1241361294  1.3590133693
## xvar        -0.0006636111  0.000600347
```

Altisteen kausaalinen vaikutus



Alkuperäinen altisteen yhteys



```
tsls_syst_rvp <- tsls(rvpfac11~syst_6_24_auc, ~SBP_UKBB_prscs_SD, data = analysisfull)
summary(tsls_syst_rvp)
```

```
##
## 2SLS Estimates
##
## Model Formula: rvpfac11 ~ syst_6_24_auc
##
## Instruments: ~SBP_UKBB_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.440732 -0.662856 -0.007088  0.000000  0.670334  3.385241
##
##           Estimate   Std. Error  t value Pr(>|t|)
## (Intercept)  0.8860824291  1.7498988117  0.50636  0.61267
## syst_6_24_auc -0.0004266307  0.0008550336 -0.49896  0.61787
##
## Residual standard error: 0.9986245 on 1767 degrees of freedom
```

```
confint(tsls_syst_rvp)
```

```
##           2.5 %      97.5 %
## (Intercept) -2.543656218  4.315821077
## syst_6_24_auc -0.002102466  0.001249204
```

```
r_syst_rvp2 <- inverse_variance_weighted(xvar= analysisfull$syst_6_24_auc,
                                         zvar = analysisfull$SBP_UKBB_prscs_SD,
                                         yvar = analysisfull$rvpfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
## IVW      0.000      0.001 -0.002, 0.001  0.618
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 100.5.
## [1] -0.0004206203
```

```
RESULTS_SystolicBP <- rows_patch(r_syst_rvp1, r_syst_rvp2, by="Test")
RESULTS_SystolicBP[3,2] <- tsls_syst_rvp$coefficients[2]
RESULTS_SystolicBP
```

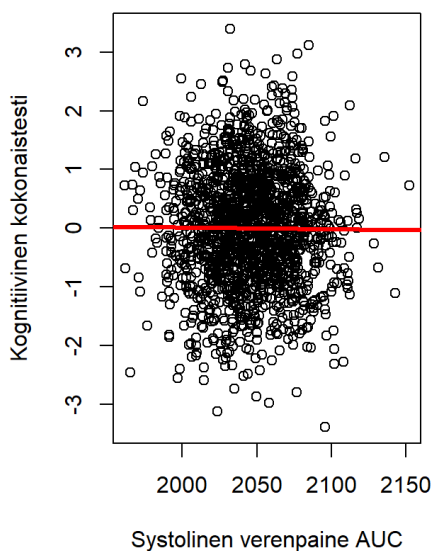
```
##           Test      Effect
## 1 Basic LM -0.0003017882
## 2 MR 2SLS1 -0.0004206203
## 3 MR 2SLS2 -0.0004266307
## 4 MR IVW -0.0004206203
## 5 MR TRIANGLE -0.0004206203
```

## Kognitiivinen kokonaissumma

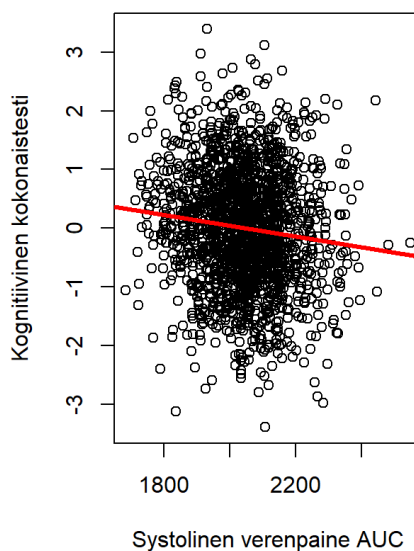
```
r_syst_cog1 <- twoSLS(xvar=analysisfull$syst_6_24_auc, xname = "Systolinen verenpaine AUC", xpred = "syst_pred",
                     zvar=analysisfull$SBP_UKBB_prscs_SD, zname = "Standardoitu systolisen verenpaineen PRSCS",
                     yvar=analysisfull$cogfac11, yname="Kognitiivinen kokonaistesti", d = analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3824  -0.6692   0.0110   0.6753   3.3818
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.6084178  1.8253012   0.333   0.739
## xpred       -0.0002947  0.0008922  -0.330   0.741
##
## (Dispersion parameter for gaussian family taken to be 1.001901)
##
## Null deviance: 1611.2 on 1609 degrees of freedom
## Residual deviance: 1611.1 on 1608 degrees of freedom
## (416 observations deleted due to missingness)
## AIC: 4576
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -2.969106778  4.185942407
## xpred       -0.002043289  0.001453972
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3334  -0.6629  -0.0031   0.6570   3.2891
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.8834467  0.3930735   4.792 1.79e-06 ***
## xvar        -0.0009216  0.0001920  -4.800 1.72e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9859583)
##
## Null deviance: 1793.5 on 1797 degrees of freedom
## Residual deviance: 1770.8 on 1796 degrees of freedom
## (228 observations deleted due to missingness)
## AIC: 5081.1
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  1.113036891  2.6538565174
## xvar        -0.001297849 -0.0005452727
```

Altisteen kausaalinen vaikutus



Alkuperäinen altisteen yhteys



```
tsls_syst_cog <- tsls(cogfac11~syst_6_24_auc, ~SBP_UKBB_prscs_SD, data = analysisfull)
summary(tsls_syst_cog)
```

```
##
## 2SLS Estimates
##
## Model Formula: cogfac11 ~ syst_6_24_auc
##
## Instruments: ~SBP_UKBB_prscs_SD
##
## Residuals:
##      Min.    1st Qu.    Median      Mean   3rd Qu.    Max.
## -3.378507 -0.665144 -0.005963  0.000000  0.666732  3.352861
##
##              Estimate   Std. Error  t value Pr(>|t|)
## (Intercept)  0.6112681962  1.8274341679  0.33450  0.73805
## syst_6_24_auc -0.0002963028  0.0008939748 -0.33144  0.74035
##
## Residual standard error: 0.9974043 on 1608 degrees of freedom
```

```
confint(tsls_syst_cog)
```

```
##              2.5 %      97.5 %
## (Intercept) -2.970436957  4.192973349
## syst_6_24_auc -0.002048461  0.001455856
```

```
r_syst_cog2 <- inverse_variance_weighted(xvar= analysisfull$syst_6_24_auc,
                                         zvar = analysisfull$SBP_UKBB_prscs_SD,
                                         yvar = analysisfull$cogfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
## IVW      0.000      0.001 -0.002, 0.001  0.741
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 100.5.
## [1] -0.0002946587
```

```
RESULTS_SystolicBP <- rows_patch(r_syst_cog1, r_syst_cog2, by="Test")
RESULTS_SystolicBP[3,2] <- tsls_syst_cog$coefficients[2]
RESULTS_SystolicBP
```

```
##      Test      Effect
## 1 Basic LM -0.0009215607
## 2 MR 2SLS1 -0.0002946587
## 3 MR 2SLS2 -0.0002963028
## 4 MR IVW -0.0002946587
## 5 MR TRIANGLE -0.0002946587
```

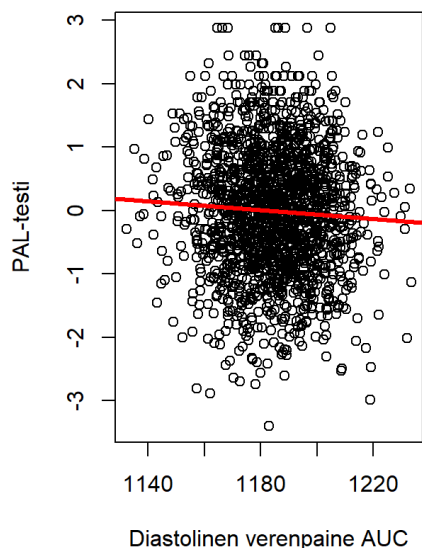
## Diastolinen verenpaine

### PAL-testi

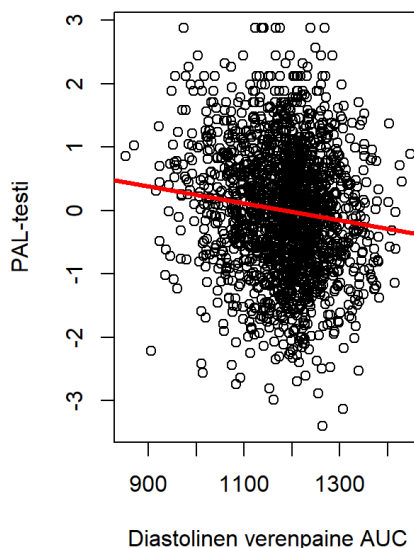
```
r_dias1 <-
twoSLS(xvar = analysisfull$dkv_6_24_auc, xname = "Diastolinen verenpaine AUC", xpred = "dias_pred",
       zvar = analysisfull$DBP_UKBB_prscs_SD, zname = "Standardoitu diastolisen verenpaineen PRSCS",
       yvar = analysisfull$palfac11, yname = "PAL-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4016  -0.6742  -0.0061   0.6693   2.9484
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.106381   1.857325   2.211  0.0272 *
## xpred        -0.003469   0.001568  -2.213  0.0271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9989636)
##
##      Null deviance: 1656.2  on 1654  degrees of freedom
## Residual deviance: 1651.3  on 1653  degrees of freedom
## (371 observations deleted due to missingness)
## AIC: 4699
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  0.466090182  7.7466715328
## xpred        -0.006541661 -0.0003961594
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2918  -0.6669  -0.0107   0.6663   2.9876
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.5990661   0.3194667   5.005 6.10e-07 ***
## xvar         -0.0013496   0.0002689  -5.019 5.71e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9848397)
##
##      Null deviance: 1842.8  on 1847  degrees of freedom
## Residual deviance: 1818.0  on 1846  degrees of freedom
## (178 observations deleted due to missingness)
## AIC: 5220.2
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  0.972922929  2.2252092495
## xvar         -0.001876698 -0.0008225296
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_dias <- tsls(palfac11~dkv_6_24_auc, ~DBP_UKBB_prscs_SD, data = analysisfull)
summary(tsls_dias)
```

```
##
## 2SLS Estimates
##
## Model Formula: palfac11 ~ dkv_6_24_auc
##
## Instruments: ~DBP_UKBB_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.152788 -0.678890 -0.007143  0.000000  0.687379  3.158066
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.986694892  1.815391032  2.19605 0.028227 *
## dkv_6_24_auc -0.003366792  0.001531861 -2.19784 0.028098 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.0062174 on 1653 degrees of freedom
```

```
confint(tsls_dias)
```

```
##              2.5 %      97.5 %
## (Intercept)  0.428593851  7.5447959337
## dkv_6_24_auc -0.006369185 -0.0003643989
```

```
r_dias2 <-
inverse_variance_weighted(analysisfull$dkv_6_24_auc, analysisfull$DBP_UKBB_prscs_SD,
                           analysisfull$palfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
##   IVW   -0.003    0.002 -0.007, 0.000  0.027
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 62.4.
## [1] -0.00346891
```

```
RESULTS_DiastolicBP <- rows_patch(r_dias1, r_dias2, by="Test")
RESULTS_DiastolicBP[3,2] <- tsls_dias$coefficients[2]
RESULTS_DiastolicBP
```

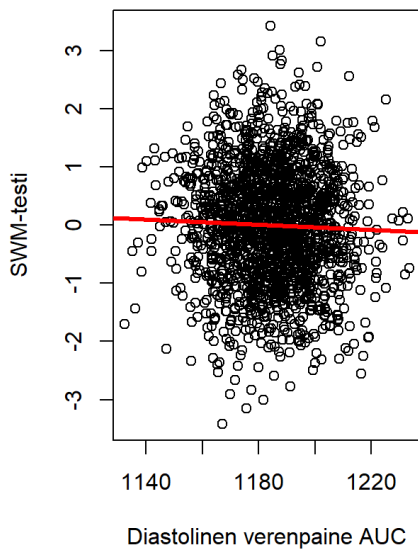
```
##           Test           Effect
## 1 Basic LM -0.001349614
## 2 MR 2SLS1 -0.003468910
## 3 MR 2SLS2 -0.003366792
## 4 MR IVW -0.003468910
## 5 MR TRIANGLE -0.003468910
```

## SWM-testi

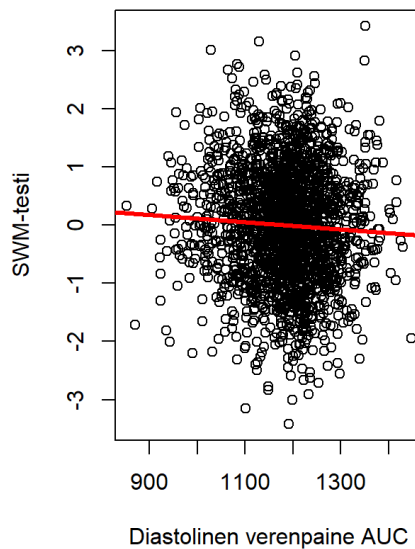
```
r_dias_swm1 <- twoSLS(xvar=analysisfull$dkv_6_24_auc, xname = "Diastolinen verenpaine AUC", xpred= "dias_pred",
  zvar=analysisfull$DBP_UKBB_prscs_SD, zname = "Standardoitu diastolisen verenpaineen PRSCS",
  yvar=analysisfull$swmfac11 , yname="SWM-testi", d = analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4611  -0.6662  -0.0095   0.6653   3.4199
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.591222   1.796334   1.443   0.149
## xpred        -0.002186   0.001516  -1.442   0.149
##
## (Dispersion parameter for gaussian family taken to be 0.9949737)
##
## Null deviance: 1792  on 1800  degrees of freedom
## Residual deviance: 1790  on 1799  degrees of freedom
## (225 observations deleted due to missingness)
## AIC: 5105.9
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept) -0.929528452  6.1119717180
## xpred        -0.005157713  0.0007847384
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4178  -0.6739   0.0041   0.6843   3.5275
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.7455089  0.3122863   2.387  0.0171 *
## xvar        -0.0006295  0.0002630  -2.393  0.0168 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9958887)
##
## Null deviance: 2006.4  on 2010  degrees of freedom
## Residual deviance: 2000.7  on 2009  degrees of freedom
## (15 observations deleted due to missingness)
## AIC: 5702.7
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept)  0.133439019  1.3575787944
## xvar        -0.001144984 -0.0001139857
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_dias_swm <- tsls(swmfac11~dkv_6_24_auc, ~DBP_UKBB_prscs_SD, data = analysisfull)
summary(tsls_dias_swm)
```

```
##
## 2SLS Estimates
##
## Model Formula: swmfac11 ~ dkv_6_24_auc
##
## Instruments: ~DBP_UKBB_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.408514 -0.683892  0.008374  0.000000  0.666482  3.788494
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.602603127  1.815809071  1.43330  0.15195
## dkv_6_24_auc -0.002196431  0.001532631 -1.43311  0.15200
##
## Residual standard error: 1.0038883 on 1799 degrees of freedom
```

```
confint(tsls_dias_swm)
```

```
##           2.5 %       97.5 %
## (Intercept) -0.956317255  6.1615235088
## dkv_6_24_auc -0.005200333  0.0008074714
```

```
r_dias_swm2 <- inverse_variance_weighted(xvar= analysisfull$dkv_6_24_auc,
                                         zvar = analysisfull$DBP_UKBB_prscs_SD,
                                         yvar = analysisfull$swmfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
## IVW   -0.002    0.002 -0.005, 0.001  0.149
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 62.4.
## [1] -0.002186487
```

```
RESULTS_DiastolicBP <- rows_patch(r_dias_swm1, r_dias_swm2, by="Test")
RESULTS_DiastolicBP[3,2] <- tsls_dias_swm$coefficients[2]
RESULTS_DiastolicBP
```

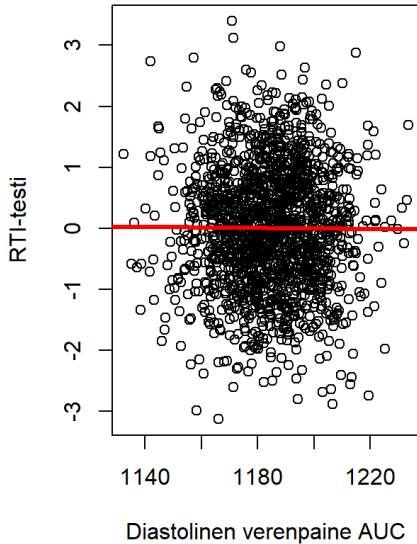
```
##          Test          Effect
## 1  Basic LM -0.0006294849
## 2  MR 2SLS1 -0.0021864874
## 3  MR 2SLS2 -0.0021964306
## 4  MR IVW -0.0021864874
## 5 MR TRIANGLE -0.0021864874
```

## RTI-testi

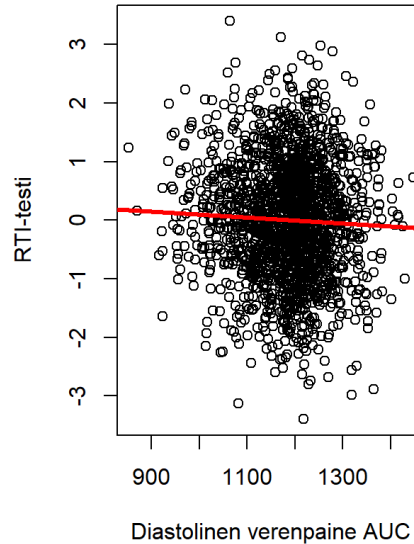
```
r_dias_rti1 <- twoSLS(xvar=analysisfull$dkv_6_24_auc, xname = "Diastolinen verenpaine AUC", xpred= "dias_pred",
  zvar=analysisfull$DBP_UKBB_prscs_SD, zname = "Standardoitu diastolisen verenpaineen PRSCS",
  yvar=analysisfull$rtifac11 , yname="RTI-testi", d = analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1386  -0.6801  -0.0018   0.6610   3.3826
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.510176   1.872575   0.272   0.785
## xpred        -0.000425   0.001581  -0.269   0.788
##
## (Dispersion parameter for gaussian family taken to be 1.000716)
##
## Null deviance: 1629.2 on 1629 degrees of freedom
## Residual deviance: 1629.2 on 1628 degrees of freedom
## (396 observations deleted due to missingness)
## AIC: 4630.9
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept) -3.16000405  4.180355892
## xpred        -0.00352302  0.002673046
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3784  -0.6804   0.0047   0.6777   3.3338
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.6023137   0.3242447   1.858   0.0634 .
## xvar        -0.0005084   0.0002730  -1.862   0.0627 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9967225)
##
## Null deviance: 1817.5 on 1821 degrees of freedom
## Residual deviance: 1814.0 on 1820 degrees of freedom
## (204 observations deleted due to missingness)
## AIC: 5168.6
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept) -0.033194109  1.237822e+00
## xvar        -0.001043403  2.662057e-05
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_dias_rti <- tsls(rtifac11~dkv_6_24_auc, ~DBP_UKBB_prscs_SD, data = analysisfull)
summary(tsls_dias_rti)
```

```
##
## 2SLS Estimates
##
## Model Formula: rtifac11 ~ dkv_6_24_auc
##
## Instruments: ~DBP_UKBB_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.174153 -0.684356  0.007914  0.000000  0.672639  3.337841
##
##              Estimate   Std. Error  t value Pr(>|t|)
## (Intercept)  0.5023638749  1.8414765918  0.27280  0.78504
## dkv_6_24_auc -0.0004182757  0.0015539691 -0.26917  0.78784
##
## Residual standard error: 0.9992477 on 1628 degrees of freedom
```

```
confint(tsls_dias_rti)
```

```
##              2.5 %      97.5 %
## (Intercept) -3.106863923  4.111591673
## dkv_6_24_auc -0.003463999  0.002627448
```

```
r_dias_rti2 <- inverse_variance_weighted(xvar= analysisfull$dkv_6_24_auc,
                                         zvar = analysisfull$DBP_UKBB_prscs_SD,
                                         yvar = analysisfull$rtifac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
## IVW      0.000      0.002 -0.004, 0.003  0.788
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 62.4.
## [1] -0.0004249873
```

```
RESULTS_DiastolicBP <- rows_patch(r_dias_rti1, r_dias_rti2, by="Test")
RESULTS_DiastolicBP[3,2] <- tsls_dias_rti$coefficients[2]
RESULTS_DiastolicBP
```

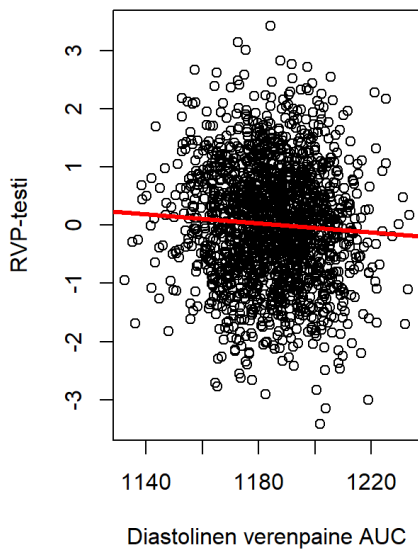
```
##          Test          Effect
## 1 Basic LM -0.0005083912
## 2 MR 2SLS1 -0.0004249873
## 3 MR 2SLS2 -0.0004182757
## 4 MR IVW -0.0004249873
## 5 MR TRIANGLE -0.0004249873
```

## RVP-testi

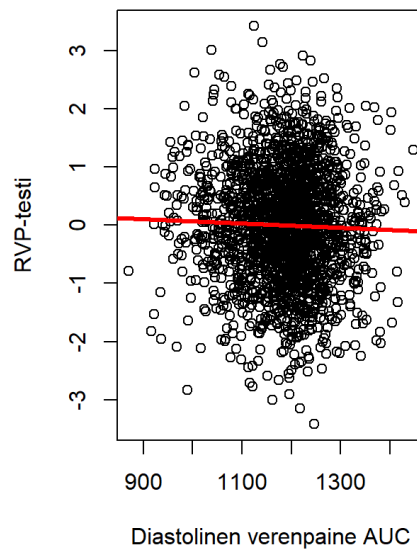
```
r_dias_rvp1 <- twoSLS(xvar=analysisfull$dkv_6_24_auc, xname = "Diastolinen verenpaine AUC", xpred= "dias_pred",
  zvar=analysisfull$DBP_UKBB_prscs_SD, zname = "Standardoitu diastolisen verenpaineen PRSCS",
  yvar=analysisfull$rvpfac11, yname="RVP-testi", d = analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3645  -0.6757  -0.0011   0.6667   3.4019
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.602769   1.813038   2.539  0.0112 *
## xpred        -0.003874   0.001530  -2.532  0.0114 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.994496)
##
## Null deviance: 1763.6 on 1768 degrees of freedom
## Residual deviance: 1757.3 on 1767 degrees of freedom
## (257 observations deleted due to missingness)
## AIC: 5014.4
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept)  1.049279954  8.1562576554
## xpred        -0.006872739 -0.0008748679
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3934  -0.6739  -0.0054   0.6767   3.3945
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.4509688  0.3181432   1.418  0.156
## xvar        -0.0003808  0.0002680  -1.421  0.155
##
## (Dispersion parameter for gaussian family taken to be 0.9976883)
##
## Null deviance: 1970.5 on 1974 degrees of freedom
## Residual deviance: 1968.4 on 1973 degrees of freedom
## (51 observations deleted due to missingness)
## AIC: 5604.2
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept) -0.1725805093  1.074518088
## xvar        -0.0009060049  0.000144411
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_dias_rvp <- tsls(rvpfac11~dkv_6_24_auc, ~DBP_UKBB_prscs_SD, data = analysisfull)
summary(tsls_dias_rvp)
```

```
##
## 2SLS Estimates
##
## Model Formula: rvpfac11 ~ dkv_6_24_auc
##
## Instruments: ~DBP_UKBB_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.642602 -0.724722  0.002462  0.000000  0.690714  3.157195
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.885662220  2.016227932  2.42317 0.015486 *
## dkv_6_24_auc -0.004112858  0.001701714 -2.41689 0.015754 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.0446308 on 1767 degrees of freedom
```

```
confint(tsls_dias_rvp)
```

```
##              2.5 %      97.5 %
## (Intercept)  0.933928088  8.837396324
## dkv_6_24_auc -0.007448155 -0.0007775605
```

```
r_dias_rvp2 <- inverse_variance_weighted(xvar= analysisfull$dkv_6_24_auc,
                                         zvar = analysisfull$DBP_UKBB_prscs_SD,
                                         yvar = analysisfull$rvpfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
##   IVW   -0.004    0.002 -0.007, -0.001  0.011
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 62.4.
## [1] -0.003873803
```

```
RESULTS_DiastolicBP <- rows_patch(r_dias_rvp1, r_dias_rvp2, by="Test")
RESULTS_DiastolicBP[3,2] <- tsls_dias_rvp$coefficients[2]
RESULTS_DiastolicBP
```

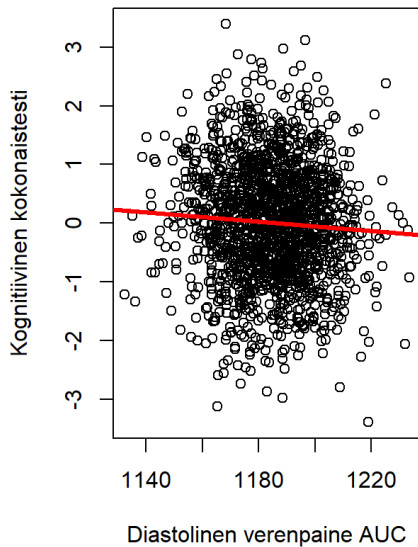
```
##           Test           Effect
## 1 Basic LM -0.0003807969
## 2 MR 2SLS1 -0.0038738034
## 3 MR 2SLS2 -0.0041128578
## 4 MR IVW -0.0038738034
## 5 MR TRIANGLE -0.0038738034
```

## Kognitiivinen kokonaissumma

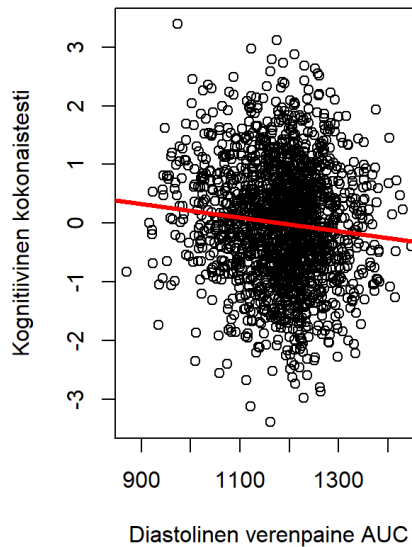
```
r_dias_cog1 <- twoSLS(xvar=analysisfull$dkv_6_24_auc, xname = "Diastolinen verenpaine AUC", xpred= "dias_pred",
  zvar=analysisfull$DBP_UKBB_prscs_SD, zname = "Standardoitu diastolisen verenpaineen PRSCS",
  yvar=analysisfull$cogfac11, yname="Kognitiivinen kokonaistesti", d = analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2607  -0.6672   0.0112   0.6746   3.3216
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.703895   1.885201   2.495  0.0127 *
## xpred        -0.003966   0.001591  -2.492  0.0128 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9981132)
##
## Null deviance: 1611.2 on 1609 degrees of freedom
## Residual deviance: 1605.0 on 1608 degrees of freedom
## (416 observations deleted due to missingness)
## AIC: 4569.9
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %          97.5 %
## (Intercept)  1.008969175  8.3988212271
## xpred        -0.007085411 -0.0008473079
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4183  -0.6673  -0.0153   0.6761   3.1459
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.382474   0.327759   4.218 2.59e-05 ***
## xvar        -0.001167   0.000276  -4.229 2.47e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9887621)
##
## Null deviance: 1793.5 on 1797 degrees of freedom
## Residual deviance: 1775.8 on 1796 degrees of freedom
## (228 observations deleted due to missingness)
## AIC: 5086.2
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %          97.5 %
## (Intercept)  0.740079114  2.0248692393
## xvar        -0.001707954 -0.0006261461
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_dias_cog <- tsls(cogfac11~dkv_6_24_auc, ~DBP_UKBB_prscs_SD, data = analysisfull)
summary(tsls_dias_cog)
```

```
##
## 2SLS Estimates
##
## Model Formula: cogfac11 ~ dkv_6_24_auc
##
## Instruments: ~DBP_UKBB_prscs_SD
##
## Residuals:
##   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.49318 -0.67769 -0.01434  0.00000  0.72499  3.07312
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.902687817  2.010541428  2.43849 0.014856 *
## dkv_6_24_auc -0.004132887  0.001696667 -2.43589 0.014963 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.0222346 on 1608 degrees of freedom
```

```
confint(tsls_dias_cog)
```

```
##           2.5 %      97.5 %
## (Intercept)  0.962099028  8.843276605
## dkv_6_24_auc -0.007458294 -0.000807481
```

```
r_dias_cog2 <- inverse_variance_weighted(xvar= analysisfull$dkv_6_24_auc,
                                         zvar = analysisfull$DBP_UKBB_prscs_SD,
                                         yvar = analysisfull$cogfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
##   IVW   -0.004    0.002 -0.007, -0.001  0.013
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 62.4.
## [1] -0.003966359
```

```
RESULTS_DiastolicBP <- rows_patch(r_dias_cog1, r_dias_cog2, by="Test")
RESULTS_DiastolicBP[3,2] <- tsls_dias_cog$coefficients[2]
RESULTS_DiastolicBP
```

```
##           Test           Effect
## 1 Basic LM -0.001167050
## 2 MR 2SLS1 -0.003966359
## 3 MR 2SLS2 -0.004132887
## 4 MR IVW -0.003966359
## 5 MR TRIANGLE -0.003966359
```

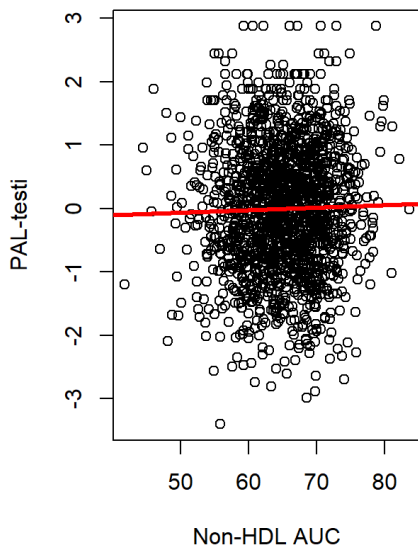
## Non-HDL-kolesteroli

### PAL-testi

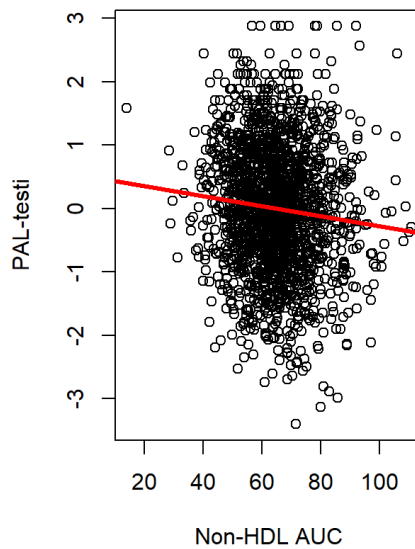
```
r_nonhdl1 <-
twoSLS(xvar = analysisfull$nonhdl_6_24_auc, xname = "Non-HDL AUC", xpred = "nonhdl_pred",
       zvar = analysisfull$nonHDL_GLGC_prscs_SD, zname = "Standardized NonHDL PRSCS",
       yvar = analysisfull$palfac11, yname = "PAL-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3579  -0.6763   0.0042   0.6759   2.9016
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.273661   0.276328  -0.990   0.322
## xpred         0.004166   0.004234   0.984   0.325
##
## (Dispersion parameter for gaussian family taken to be 1.001336)
##
## Null deviance: 1656.2  on 1654  degrees of freedom
## Residual deviance: 1655.2  on 1653  degrees of freedom
## (371 observations deleted due to missingness)
## AIC: 4702.9
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.815253211  0.26793210
## xpred        -0.004133261  0.01246493
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3464  -0.6799   0.0092   0.6572   3.0864
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.507903   0.128521   3.952 8.05e-05 ***
## xvar        -0.007820   0.001946  -4.018 6.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9896234)
##
## Null deviance: 1842.8  on 1847  degrees of freedom
## Residual deviance: 1826.8  on 1846  degrees of freedom
## (178 observations deleted due to missingness)
## AIC: 5229.1
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  0.25600574  0.759799934
## xvar        -0.01163492 -0.004004994
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_nonhdl <- tsls(palfac11~nonhdl_6_24_auc, ~nonHDL_GLGC_prscs_SD, data = analysisfull)
summary(tsls_nonhdl)
```

```
##
## 2SLS Estimates
##
## Model Formula: palfac11 ~ nonhdl_6_24_auc
##
## Instruments: ~nonHDL_GLGC_prscs_SD
##
## Residuals:
##   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.424846 -0.668597  0.006148  0.000000  0.675897  2.912683
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.278037233  0.282409612 -0.98452  0.32501
## nonhdl_6_24_auc  0.004239488  0.004334509  0.97808  0.32818
##
## Residual standard error: 1.0065504 on 1653 degrees of freedom
```

```
confint(tsls_nonhdl)
```

```
##           2.5 %   97.5 %
## (Intercept) -0.831549901 0.27547544
## nonhdl_6_24_auc -0.004255993 0.01273497
```

```
r_nonhdl2 <-
inverse_variance_weighted(analysisfull$nonhdl_6_24_auc, analysisfull$nonHDL_GLGC_prscs_SD,
                           analysisfull$palfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI p-value
## IVW 0.004 0.004 -0.004, 0.012 0.325
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 558.4.
## [1] 0.004165832
```

```
RESULTS_NonHDL <- rows_patch(r_nonhdl1, r_nonhdl2, by="Test")
RESULTS_NonHDL[3,2] <- tsls_nonhdl$coefficients[2]
RESULTS_NonHDL
```

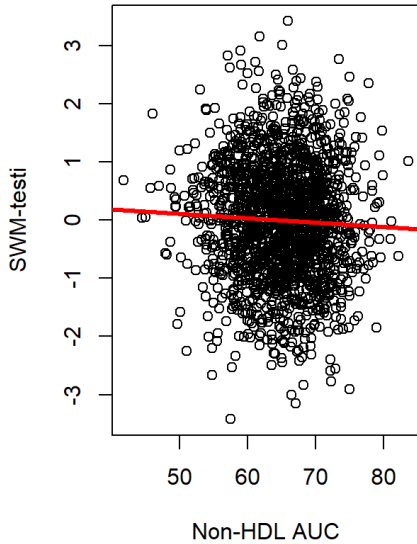
```
##          Test      Effect
## 1  Basic LM -0.007819955
## 2  MR 2SLS1  0.004165832
## 3  MR 2SLS2  0.004239488
## 4  MR IVW   0.004165832
## 5  MR TRIANGLE 0.004165832
```

## SWM-testi

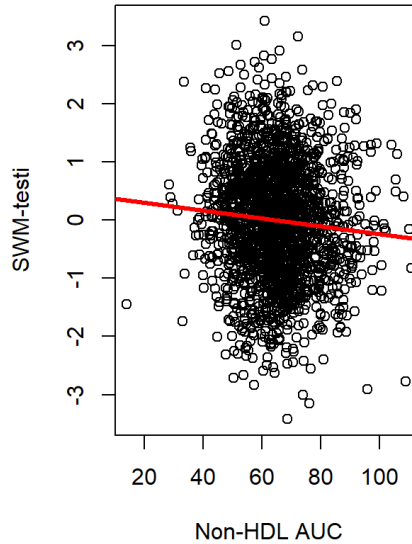
```
r_nonhdl1_swm <-
  twoSLS(xvar = analysisfull$nonhdl_6_24_auc, xname = "Non-HDL AUC", xpred = "nonhdl_pred",
        zvar = analysisfull$nonHDL_GLGC_prscs_SD, zname = "Standardized NonHDL PRSCS",
        yvar = analysisfull$swmfac11, yname = "SWM-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4791  -0.6666  -0.0062   0.6666   3.4282
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.484249   0.265269   1.826   0.0681 .
## xpred        -0.007433   0.004061  -1.831   0.0673 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9942722)
##
## Null deviance: 1792.0  on 1800  degrees of freedom
## Residual deviance: 1788.7  on 1799  degrees of freedom
## (225 observations deleted due to missingness)
## AIC: 5104.7
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.03566776  1.0041661556
## xpred        -0.01539243  0.0005254865
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3968  -0.6696   0.0027   0.6656   3.3949
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.444013   0.123243   3.603 0.000323 ***
## xvar         -0.006821   0.001862  -3.663 0.000256 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9921033)
##
## Null deviance: 2006.4  on 2010  degrees of freedom
## Residual deviance: 1993.1  on 2009  degrees of freedom
## (15 observations deleted due to missingness)
## AIC: 5695
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  0.20246066  0.685564563
## xvar         -0.01047052 -0.003170851
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_nonhdl_swm <- tsls(swmfac11~nonhdl_6_24_auc, ~nonHDL_GLGC_prscs_SD, data = analysisfull)
summary(tsls_nonhdl_swm)
```

```
##
## 2SLS Estimates
##
## Model Formula: swmfac11 ~ nonhdl_6_24_auc
##
## Instruments: ~nonHDL_GLGC_prscs_SD
##
## Residuals:
##   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.39475 -0.66330  0.00192  0.00000  0.66624  3.39236
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.482948726  0.263919470  1.82991 0.067429 .
## nonhdl_6_24_auc -0.007416221  0.004041520 -1.83501 0.066670 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9947141 on 1799 degrees of freedom
```

```
confint(tsls_nonhdl_swm)
```

```
##           2.5 %      97.5 %
## (Intercept)  -0.03432393  1.0002213817
## nonhdl_6_24_auc -0.01533745  0.0005050121
```

```
r_nonhdl2_swm <-
  inverse_variance_weighted(analysisfull$nonhdl_6_24_auc, analysisfull$nonHDL_GLGC_prscs_SD,
    analysisfull$swmfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
##   IVW   -0.007    0.004 -0.015, 0.001  0.067
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 558.4.
## [1] -0.007433474
```

```
RESULTS_NonHDL <- rows_patch(r_nonhdl1_swm, r_nonhdl2_swm, by="Test")
RESULTS_NonHDL[3,2] <- tsls_nonhdl_swm$coefficients[2]
RESULTS_NonHDL
```

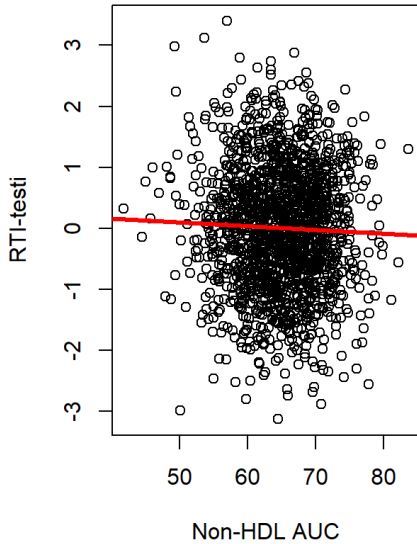
```
##           Test      Effect
## 1   Basic LM -0.006820685
## 2     MR 2SLS1 -0.007433474
## 3     MR 2SLS2 -0.007416221
## 4       MR IVW -0.007433474
## 5 MR TRIANGLE -0.007433474
```

## RTI-testi

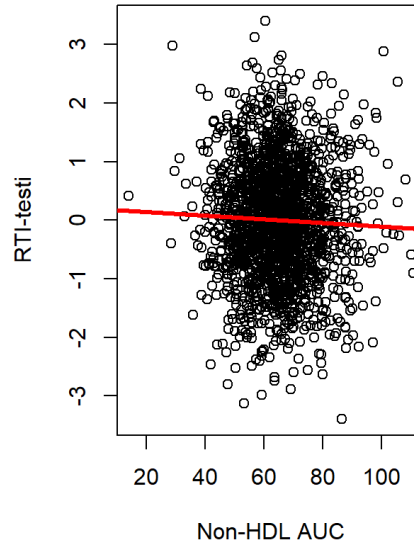
```
r_nonhdl1_rti <-
  twoSLS(xvar = analysisfull$nonhdl_6_24_auc, xname = "Non-HDL AUC", xpred = "nonhdl_pred",
        zvar = analysisfull$nonHDL_GLGC_prscs_SD, zname = "Standardized NonHDL PRSCS",
        yvar = analysisfull$rtifac11, yname = "RTI-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1337  -0.6745  -0.0017   0.6622   3.3389
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.409127   0.277935   1.472   0.141
## xpred        -0.006192   0.004260  -1.454   0.146
##
## (Dispersion parameter for gaussian family taken to be 0.9994632)
##
## Null deviance: 1629.2 on 1629 degrees of freedom
## Residual deviance: 1627.1 on 1628 degrees of freedom
## (396 observations deleted due to missingness)
## AIC: 4628.9
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.13561546  0.953869347
## xpred        -0.01454183  0.002157443
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3291  -0.6791   0.0026   0.6676   3.3815
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.198054   0.130046   1.523   0.128
## xvar        -0.003053   0.001972  -1.548   0.122
##
## (Dispersion parameter for gaussian family taken to be 0.9973087)
##
## Null deviance: 1817.5 on 1821 degrees of freedom
## Residual deviance: 1815.1 on 1820 degrees of freedom
## (204 observations deleted due to missingness)
## AIC: 5169.7
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.05683182  0.4529402149
## xvar        -0.00691884  0.0008120585
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_nonhdl_rti <- tsls(rtifac11~nonhdl_6_24_auc, ~nonHDL_GLGC_prscs_SD, data = analysisfull)
summary(tsls_nonhdl_rti)
```

```
##
## 2SLS Estimates
##
## Model Formula: rtifac11 ~ nonhdl_6_24_auc
##
## Instruments: ~nonHDL_GLGC_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.203427 -0.690189  0.003123  0.000000  0.673607  3.360469
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.417564009  0.284037170  1.47010  0.14173
## nonhdl_6_24_auc -0.006339135  0.004366114 -1.45189  0.14672
##
## Residual standard error: 1.0008604 on 1628 degrees of freedom
```

```
confint(tsls_nonhdl_rti)
```

```
##              2.5 %    97.5 %
## (Intercept)  -0.13913861 0.97426663
## nonhdl_6_24_auc -0.01489656 0.00221829
```

```
r_nonhdl2_rti <-
  inverse_variance_weighted(analysisfull$nonhdl_6_24_auc, analysisfull$nonHDL_GLGC_prscs_SD,
    analysisfull$rtifac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
## IVW   -0.006    0.004 -0.015, 0.002  0.146
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 558.4.
## [1] -0.006192196
```

```
RESULTS_NonHDL <- rows_patch(r_nonhdl1_rti, r_nonhdl2_rti, by="Test")
RESULTS_NonHDL[3,2] <- tsls_nonhdl_rti$coefficients[2]
RESULTS_NonHDL
```

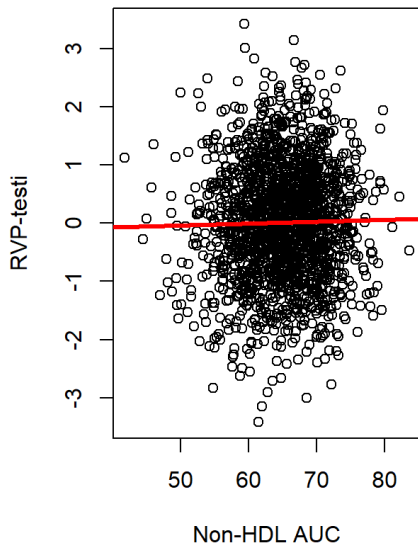
```
##          Test      Effect
## 1  Basic LM -0.003053391
## 2  MR 2SLS1 -0.006192196
## 3  MR 2SLS2 -0.006339135
## 4  MR IVW  -0.006192196
## 5  MR TRIANGLE -0.006192196
```

## RVP-testi

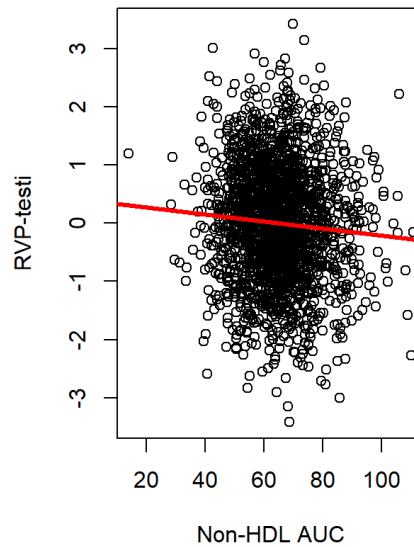
```
r_nonhdl1_rvp <-
  twoSLS(xvar = analysisfull$nonhdl_6_24_auc, xname = "Non-HDL AUC", xpred = "nonhdl_pred",
        zvar = analysisfull$nonHDL_GLCG_prscs_SD, zname = "Standardized NonHDL PRSCS",
        yvar = analysisfull$rvpfac11, yname = "RVP-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4187  -0.6665  -0.0078   0.6728   3.4219
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.193156   0.268046  -0.721   0.471
## xpred        0.003170   0.004105   0.772   0.440
##
## (Dispersion parameter for gaussian family taken to be 0.9977667)
##
## Null deviance: 1763.6  on 1768  degrees of freedom
## Residual deviance: 1763.1  on 1767  degrees of freedom
## (257 observations deleted due to missingness)
## AIC: 5020.2
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %    97.5 %
## (Intercept) -0.718516207 0.33220330
## xpred       -0.004876157 0.01121706
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3949  -0.6591   0.0034   0.6782   3.4464
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.392088   0.124726   3.144 0.00169 **
## xvar        -0.006027   0.001886  -3.196 0.00142 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9935666)
##
## Null deviance: 1970.5  on 1974  degrees of freedom
## Residual deviance: 1960.3  on 1973  degrees of freedom
## (51 observations deleted due to missingness)
## AIC: 5596.1
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %    97.5 %
## (Intercept)  0.147629723 0.636546576
## xvar        -0.009722702 -0.002330377
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_nonhdl_rvp <- tsls(rvpfac11~nonhdl_6_24_auc, ~nonHDL_GLGC_prscs_SD, data = analysisfull)
summary(tsls_nonhdl_rvp)
```

```
##
## 2SLS Estimates
##
## Model Formula: rvpfac11 ~ nonhdl_6_24_auc
##
## Instruments: ~nonHDL_GLGC_prscs_SD
##
## Residuals:
##   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.442068 -0.674253 -0.002318  0.000000  0.669117  3.388525
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.193095197  0.268795503  -0.71837  0.47262
## nonhdl_6_24_auc  0.003171197  0.004119155  0.76987  0.44148
##
## Residual standard error: 1.0019726 on 1767 degrees of freedom
```

```
confint(tsls_nonhdl_rvp)
```

```
##           2.5 %    97.5 %
## (Intercept)  -0.719924702  0.33373431
## nonhdl_6_24_auc -0.004902199  0.01124459
```

```
r_nonhdl2_rvp <-
  inverse_variance_weighted(analysisfull$nonhdl_6_24_auc, analysisfull$nonHDL_GLGC_prscs_SD,
    analysisfull$rvpfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI p-value
## IVW 0.003 0.004 -0.005, 0.011 0.440
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 558.4.
## [1] 0.003170452
```

```
RESULTS_NonHDL <- rows_patch(r_nonhdl1_rvp, r_nonhdl2_rvp, by="Test")
RESULTS_NonHDL[3,2] <- tsls_nonhdl_rvp$coefficients[2]
RESULTS_NonHDL
```

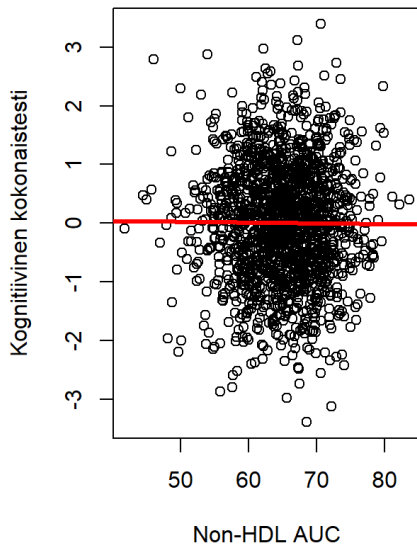
```
##          Test      Effect
## 1   Basic LM -0.006026539
## 2   MR 2SLS1  0.003170452
## 3   MR 2SLS2  0.003171197
## 4     MR IVW  0.003170452
## 5 MR TRIANGLE 0.003170452
```

## Kognitiivinen kokonaissumma

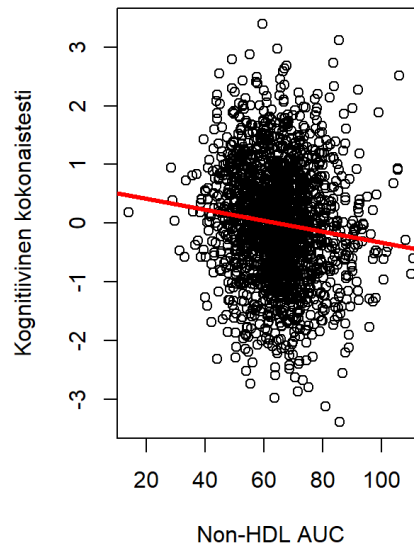
```
r_nonhdl1_cog <-
  twoSLS(xvar = analysisfull$nonhdl_6_24_auc, xname = "Non-HDL AUC", xpred = "nonhdl_pred",
        zvar = analysisfull$nonHDL_GLGC_prscs_SD, zname = "Standardized NonHDL PRSCS",
        yvar = analysisfull$cogfac11, yname = "Kognitiivinen kokonaistesti", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3915  -0.6654   0.0073   0.6748   3.3948
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.108004   0.279866   0.386   0.700
## xpred        -0.001576   0.004291  -0.367   0.713
##
## (Dispersion parameter for gaussian family taken to be 1.001885)
##
## Null deviance: 1611.2 on 1609 degrees of freedom
## Residual deviance: 1611.0 on 1608 degrees of freedom
## (416 observations deleted due to missingness)
## AIC: 4576
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.440523677  0.656531660
## xpred        -0.009985508  0.006833931
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1975  -0.6567   0.0115   0.6573   3.3422
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.602170   0.130448   4.616 4.19e-06 ***
## xvar         -0.009282   0.001978  -4.692 2.90e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9865127)
##
## Null deviance: 1793.5 on 1797 degrees of freedom
## Residual deviance: 1771.8 on 1796 degrees of freedom
## (228 observations deleted due to missingness)
## AIC: 5082.1
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  0.34649605   0.857844227
## xvar         -0.01315904  -0.005405063
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_nonhdl_cog <- tsls(cogfac11~nonhdl_6_24_auc, ~nonHDL_GLGC_prscs_SD, data = analysisfull)
summary(tsls_nonhdl_cog)
```

```
##
## 2SLS Estimates
##
## Model Formula: cogfac11 ~ nonhdl_6_24_auc
##
## Instruments: ~nonHDL_GLGC_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.363386 -0.670637  0.005358  0.000000  0.675611  3.377389
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.110026291  0.284870212  0.38623  0.69938
## nonhdl_6_24_auc -0.001610408  0.004377626 -0.36787  0.71302
##
## Residual standard error: 0.9992547 on 1608 degrees of freedom
```

```
confint(tsls_nonhdl_cog)
```

```
##              2.5 %      97.5 %
## (Intercept)  -0.4483091  0.668361646
## nonhdl_6_24_auc -0.0101904  0.006969582
```

```
r_nonhdl2_cog <-
  inverse_variance_weighted(analysisfull$nonhdl_6_24_auc, analysisfull$nonHDL_GLGC_prscs_SD,
    analysisfull$cogfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
## IVW   -0.002    0.004 -0.010, 0.007  0.713
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 558.4.
## [1] -0.001575788
```

```
RESULTS_NonHDL <- rows_patch(r_nonhdl1_cog, r_nonhdl2_cog, by="Test")
RESULTS_NonHDL[3,2] <- tsls_nonhdl_cog$coefficients[2]
RESULTS_NonHDL
```

```
##          Test          Effect
## 1 Basic LM -0.009282054
## 2 MR 2SLS1 -0.001575788
## 3 MR 2SLS2 -0.001610408
## 4 MR IVW -0.001575788
## 5 MR TRIANGLE -0.001575788
```

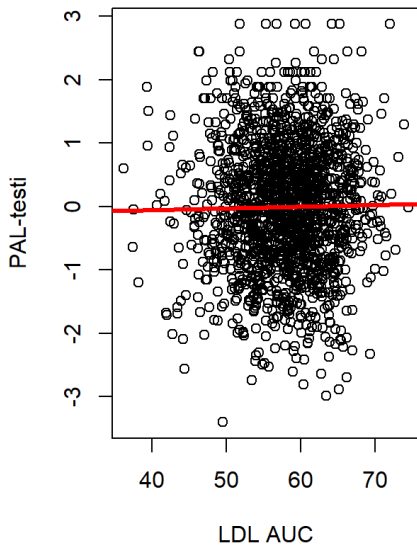
## LDL-kolesteroli

### PAL-testi

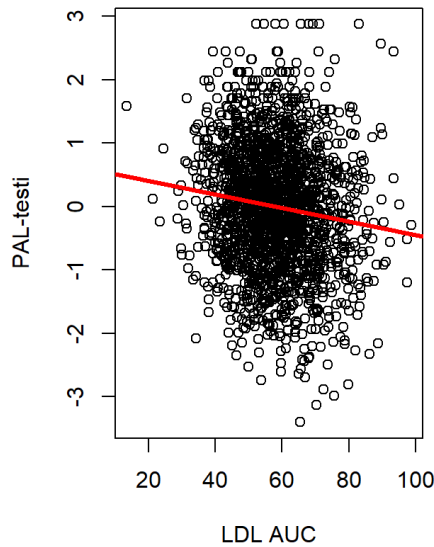
```
r_LDL1 <-
twoSLS(xvar = analysisfull$l_d1_6_24_auc, xname = "LDL AUC", xpred = "LDL_pred",
       zvar = analysisfull$LDL_GLGC_prscs_SD, zname = "Standardized LDL PRSCS",
       yvar = analysisfull$palfac11, yname = "PAL-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3726  -0.6784   0.0019   0.6762   2.8951
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.163158   0.251301  -0.649   0.516
## xpred         0.002762   0.004310   0.641   0.522
##
## (Dispersion parameter for gaussian family taken to be 1.001673)
##
## Null deviance: 1656.2 on 1654 degrees of freedom
## Residual deviance: 1655.8 on 1653 degrees of freedom
## (371 observations deleted due to missingness)
## AIC: 4703.5
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.655697982  0.32938279
## xpred        -0.005684995  0.01120897
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3195  -0.6721   0.0179   0.6737   3.1429
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.623710   0.117949   5.288 1.38e-07 ***
## xvar         -0.010739   0.001992  -5.392 7.86e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9827975)
##
## Null deviance: 1842.8 on 1847 degrees of freedom
## Residual deviance: 1814.2 on 1846 degrees of freedom
## (178 observations deleted due to missingness)
## AIC: 5216.3
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  0.39253446  0.854886370
## xvar         -0.01464317 -0.006835742
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_LDL <- tsls(palfac11~ldl_6_24_auc, ~LDL_GLGC_prscs_SD, data = analysisfull)
summary(tsls_LDL)
```

```
##
## 2SLS Estimates
##
## Model Formula: palfac11 ~ ldl_6_24_auc
##
## Instruments: ~LDL_GLGC_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.417073 -0.672701  0.004708  0.000000  0.672917  2.893770
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.164612688  0.254709535 -0.64628  0.51819
## ldl_6_24_auc  0.002789972  0.004373159  0.63798  0.52358
##
## Residual standard error: 1.0053721 on 1653 degrees of freedom
```

```
confint(tsls_LDL)
```

```
##           2.5 %    97.5 %
## (Intercept) -0.663834203  0.33460883
## ldl_6_24_auc -0.005781261  0.01136121
```

```
r_LDL2 <-
inverse_variance_weighted(analysisfull$ldl_6_24_auc, analysisfull$LDL_GLGC_prscs_SD,
                           analysisfull$palfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI p-value
## IVW 0.003 0.004 -0.006, 0.011 0.522
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 580.8.
## [1] 0.002761989
```

```
RESULTS_LDL <- rows_patch(r_LDL1, r_LDL2, by="Test")
RESULTS_LDL[3,2] <- tsls_LDL$coefficients[2]
RESULTS_LDL
```

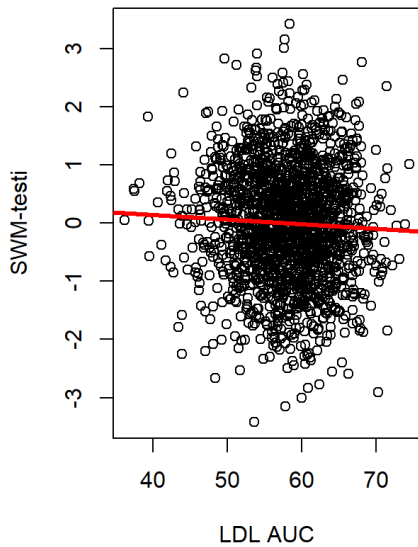
```
##          Test      Effect
## 1  Basic LM -0.010739458
## 2  MR 2SLS1  0.002761989
## 3  MR 2SLS2  0.002789972
## 4  MR IVW   0.002761989
## 5  MR TRIANGLE 0.002761989
```

## SWM-testi

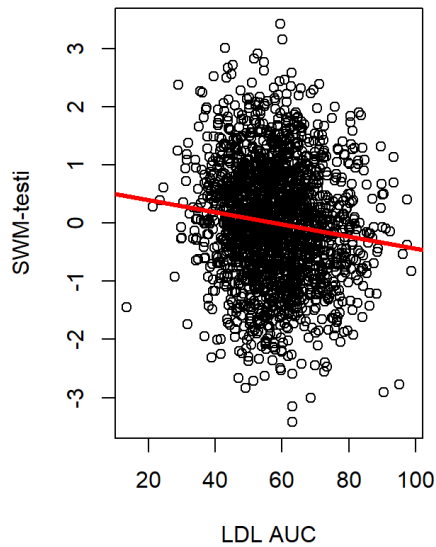
```
r_LDL1_swm <-
  twoSLS(xvar = analysisfull$ldl_6_24_auc, xname = "LDL AUC", xpred = "LDL_pred",
        zvar = analysisfull$LDL_GLCG_prscs_SD, zname = "Standardized LDL PRSCS",
        yvar = analysisfull$swmfac11, yname = "SWM-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4586  -0.6694  -0.0085   0.6681   3.4236
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.462750   0.240592   1.923  0.0546 .
## xpred        -0.007956   0.004122  -1.930  0.0537 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9940655)
##
##    Null deviance: 1792.0  on 1800  degrees of freedom
## Residual deviance: 1788.3  on 1799  degrees of freedom
## (225 observations deleted due to missingness)
## AIC: 5104.3
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.008801305  0.9343017678
## xpred        -0.016034750  0.0001224934
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3710  -0.6511  -0.0026   0.6548   3.4346
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.604331   0.113341   5.332 1.08e-07 ***
## xvar         -0.010387   0.001911  -5.437 6.09e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.984248)
##
##    Null deviance: 2006.4  on 2010  degrees of freedom
## Residual deviance: 1977.4  on 2009  degrees of freedom
## (15 observations deleted due to missingness)
## AIC: 5679
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  0.38218801  0.826474910
## xvar         -0.01413177 -0.006642379
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_LDL_swm <- tsls(swmfac11~ldl_6_24_auc, ~LDL_GLGC_prscs_SD, data = analysisfull)
summary(tsls_LDL_swm)
```

```
##
## 2SLS Estimates
##
## Model Formula: swmfac11 ~ ldl_6_24_auc
##
## Instruments: ~LDL_GLGC_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.382605 -0.656804  0.000253  0.000000  0.656432  3.432037
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.461939170  0.238821179  1.93425 0.053239 .
## ldl_6_24_auc -0.007946436  0.004093616 -1.94118 0.052393 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9914134 on 1799 degrees of freedom
```

```
confint(tsls_LDL_swm)
```

```
##           2.5 %      97.5 %
## (Intercept) -0.006141739 9.300201e-01
## ldl_6_24_auc -0.015969776 7.690327e-05
```

```
r_LDL2_swm <-
  inverse_variance_weighted(analysisfull$ldl_6_24_auc, analysisfull$LDL_GLGC_prscs_SD,
    analysisfull$swmfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
##   IVW   -0.008    0.004 -0.016, 0.000  0.054
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 580.8.
## [1] -0.007956128
```

```
RESULTS_LDL <- rows_patch(r_LDL1_swm, r_LDL2_swm, by="Test")
RESULTS_LDL[3,2] <- ts1s_LDL_swm$coefficients[2]
RESULTS_LDL
```

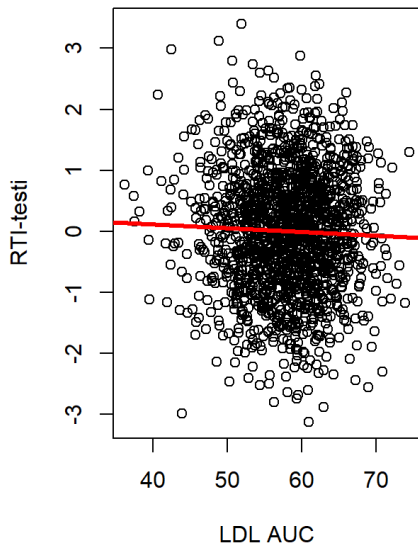
```
##           Test      Effect
## 1   Basic LM -0.010387076
## 2     MR 2SLS1 -0.007956128
## 3     MR 2SLS2 -0.007946436
## 4       MR IVW -0.007956128
## 5 MR TRIANGLE -0.007956128
```

## RTI-testi

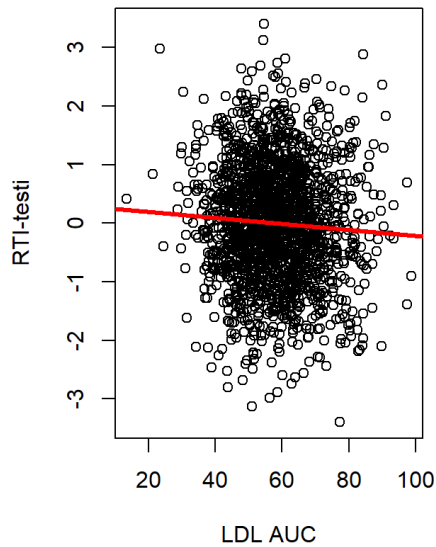
```
r_LDL1_rti <-
  twoSLS(xvar = analysisfull$ldl_6_24_auc, xname = "LDL AUC", xpred = "LDL_pred",
        zvar = analysisfull$LDL_GLGC_prscs_SD, zname = "Standardized LDL PRSCS",
        yvar = analysisfull$rtifac11, yname = "RTI-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1119  -0.6832   0.0021   0.6658   3.3497
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.375203   0.252571   1.486   0.138
## xpred       -0.006351   0.004333  -1.466   0.143
##
## (Dispersion parameter for gaussian family taken to be 0.9994411)
##
## Null deviance: 1629.2 on 1629 degrees of freedom
## Residual deviance: 1627.1 on 1628 degrees of freedom
## (396 observations deleted due to missingness)
## AIC: 4628.8
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.11982745  0.870232479
## xpred       -0.01484367  0.002140758
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2980  -0.6844   0.0058   0.6669   3.3778
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.290505   0.119610   2.429   0.0152 *
## xvar        -0.005008   0.002022  -2.477   0.0134 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9952683)
##
## Null deviance: 1817.5 on 1821 degrees of freedom
## Residual deviance: 1811.4 on 1820 degrees of freedom
## (204 observations deleted due to missingness)
## AIC: 5166
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  0.056073253  0.52493726
## xvar        -0.008971739 -0.00104458
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_LDL_rti <- tsls(rtifac11~ldl_6_24_auc, ~LDL_GLGC_prscs_SD, data = analysisfull)
summary(tsls_LDL_rti)
```

```
##
## 2SLS Estimates
##
## Model Formula: rtifac11 ~ ldl_6_24_auc
##
## Instruments: ~LDL_GLGC_prscs_SD
##
## Residuals:
##   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.17591 -0.68847  0.01208  0.00000  0.67231  3.36697
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.378619803  0.254811056  1.48588  0.13750
## ldl_6_24_auc -0.006424602  0.004381366 -1.46635  0.14275
##
## Residual standard error: 0.9994068 on 1628 degrees of freedom
```

```
confint(tsls_LDL_rti)
```

```
##           2.5 %    97.5 %
## (Intercept) -0.12080069 0.878040295
## ldl_6_24_auc -0.01501192 0.002162718
```

```
r_LDL2_rti <-
  inverse_variance_weighted(analysisfull$ldl_6_24_auc, analysisfull$LDL_GLGC_prscs_SD,
    analysisfull$rtifac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
##   IVW   -0.006    0.004 -0.015, 0.002  0.143
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 580.8.
## [1] -0.006351453
```

```
RESULTS_LDL <- rows_patch(r_LDL1_rti, r_LDL2_rti, by="Test")
RESULTS_LDL[3,2] <- tsls_LDL_rti$coefficients[2]
RESULTS_LDL
```

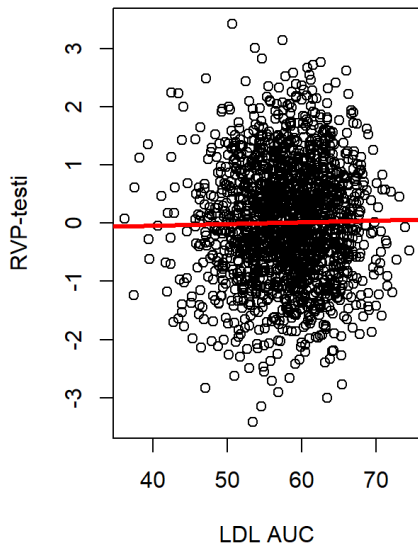
```
##          Test          Effect
## 1 Basic LM -0.005008160
## 2 MR 2SLS1 -0.006351453
## 3 MR 2SLS2 -0.006424602
## 4 MR IVW -0.006351453
## 5 MR TRIANGLE -0.006351453
```

## RVP-testi

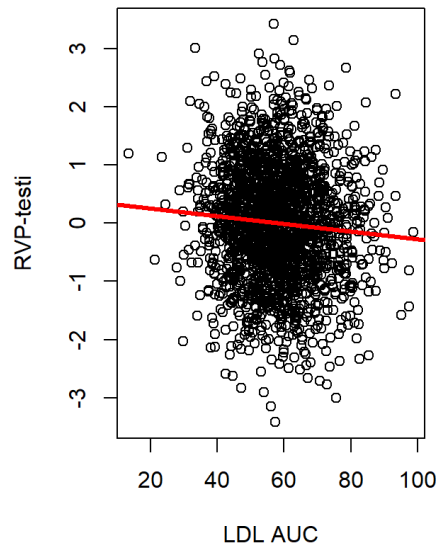
```
r_LDL1_rvp <-
  twoSLS(xvar = analysisfull$ldl_6_24_auc, xname = "LDL AUC", xpred = "LDL_pred",
        zvar = analysisfull$LDL_GLCG_prscs_SD, zname = "Standardized LDL PRSCS",
        yvar = analysisfull$rvpfac11, yname = "RVP-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4155  -0.6623  -0.0066   0.6722   3.4276
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.170753   0.242852  -0.703   0.482
## xpred        0.003166   0.004163   0.760   0.447
##
## (Dispersion parameter for gaussian family taken to be 0.997777)
##
## Null deviance: 1763.6 on 1768 degrees of freedom
## Residual deviance: 1763.1 on 1767 degrees of freedom
## (257 observations deleted due to missingness)
## AIC: 5020.3
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %    97.5 %
## (Intercept) -0.646733840  0.3052287
## xpred        -0.004993797  0.0113249
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4222  -0.6564   0.0001   0.6834   3.4098
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.381249   0.115261   3.308 0.000958 ***
## xvar        -0.006558   0.001945  -3.372 0.000760 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9929864)
##
## Null deviance: 1970.5 on 1974 degrees of freedom
## Residual deviance: 1959.2 on 1973 degrees of freedom
## (51 observations deleted due to missingness)
## AIC: 5594.9
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %    97.5 %
## (Intercept)  0.1553424  0.607156345
## xvar        -0.0103689 -0.002746137
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_LDL_rvp <- tsls(rvpfac11~ldl_6_24_auc, ~LDL_GLGC_prscs_SD, data = analysisfull)
summary(tsls_LDL_rvp)
```

```
##
## 2SLS Estimates
##
## Model Formula: rvpfac11 ~ ldl_6_24_auc
##
## Instruments: ~LDL_GLGC_prscs_SD
##
## Residuals:
##   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.428158 -0.671349 -0.001798  0.000000  0.672044  3.407235
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.171087718  0.244019795 -0.70112  0.48332
## ldl_6_24_auc  0.003173998  0.004186626  0.75813  0.44848
##
## Residual standard error: 1.0018815 on 1767 degrees of freedom
```

```
confint(tsls_LDL_rvp)
```

```
##           2.5 %    97.5 %
## (Intercept) -0.649357727  0.30718229
## ldl_6_24_auc -0.005031638  0.01137963
```

```
r_LDL2_rvp <-
  inverse_variance_weighted(analysisfull$ldl_6_24_auc, analysisfull$LDL_GLGC_prscs_SD,
    analysisfull$rvpfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI p-value
## IVW 0.003 0.004 -0.005, 0.011 0.447
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 580.8.
## [1] 0.003165552
```

```
RESULTS_LDL <- rows_patch(r_LDL1_rvp, r_LDL2_rvp, by="Test")
RESULTS_LDL[3,2] <- tsls_LDL_rvp$coefficients[2]
RESULTS_LDL
```

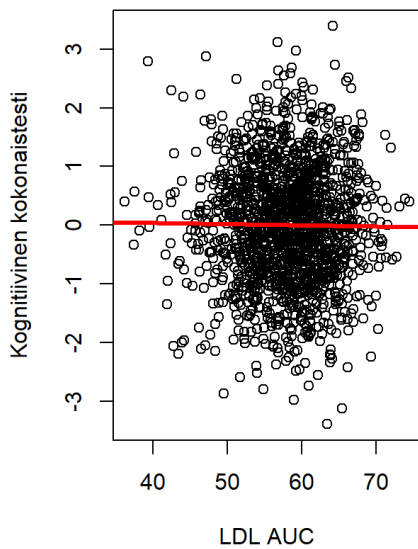
```
##          Test      Effect
## 1  Basic LM -0.006557518
## 2  MR 2SLS1  0.003165552
## 3  MR 2SLS2  0.003173998
## 4  MR IVW   0.003165552
## 5 MR TRIANGLE 0.003165552
```

## Kognitiivinen kokonaissumma

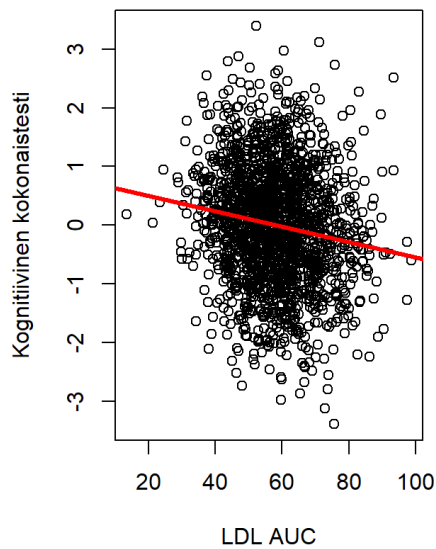
```
r_LDL1_cog <-
  twoSLS(xvar = analysisfull$ldl_6_24_auc, xname = "LDL AUC", xpred = "LDL_pred",
        zvar = analysisfull$LDL_GLGC_prscs_SD, zname = "Standardized LDL PRSCS",
        yvar = analysisfull$cogfac11, yname = "Kognitiivinen kokonaistesti", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3869  -0.6680   0.0069   0.6759   3.3976
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.115736   0.254240   0.455   0.649
## xpred        -0.001898   0.004363  -0.435   0.663
##
## (Dispersion parameter for gaussian family taken to be 1.001851)
##
## Null deviance: 1611.2 on 1609 degrees of freedom
## Residual deviance: 1611.0 on 1608 degrees of freedom
## (416 observations deleted due to missingness)
## AIC: 4576
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.38256472  0.614035963
## xpred        -0.01044883  0.006651939
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1626  -0.6613   0.0058   0.6475   3.3174
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.75658    0.11947   6.333 3.03e-10 ***
## xvar        -0.01305    0.00202  -6.457 1.37e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9759515)
##
## Null deviance: 1793.5 on 1797 degrees of freedom
## Residual deviance: 1752.8 on 1796 degrees of freedom
## (228 observations deleted due to missingness)
## AIC: 5062.7
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  0.52243150  0.990732497
## xvar        -0.01700576 -0.009085848
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_LDL_cog <- tsls(cogfac11~ldl_6_24_auc, ~LDL_GLGC_prscs_SD, data = analysisfull)
summary(tsls_LDL_cog)
```

```
##
## 2SLS Estimates
##
## Model Formula: cogfac11 ~ ldl_6_24_auc
##
## Instruments: ~LDL_GLGC_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.363137 -0.669540  0.003245  0.000000  0.672421  3.375155
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.117035037  0.256497851  0.45628  0.64825
## ldl_6_24_auc -0.001924810  0.004410836 -0.43638  0.66262
##
## Residual standard error: 0.9981489 on 1608 degrees of freedom
```

```
confint(tsls_LDL_cog)
```

```
##           2.5 %    97.5 %
## (Intercept) -0.38569151 0.61976159
## ldl_6_24_auc -0.01056989 0.00672027
```

```
r_LDL2_cog <-
  inverse_variance_weighted(analysisfull$ldl_6_24_auc, analysisfull$LDL_GLGC_prscs_SD,
    analysisfull$cogfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI p-value
## IVW -0.002 0.004 -0.010, 0.007 0.663
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 580.8.
## [1] -0.001898446
```

```
RESULTS_LDL <- rows_patch(r_LDL1_cog, r_LDL2_cog, by="Test")
RESULTS_LDL[3,2] <- tsls_LDL_cog$coefficients[2]
RESULTS_LDL
```

```
##          Test          Effect
## 1 Basic LM -0.013045804
## 2 MR 2SLS1 -0.001898446
## 3 MR 2SLS2 -0.001924810
## 4 MR IVW -0.001898446
## 5 MR TRIANGLE -0.001898446
```

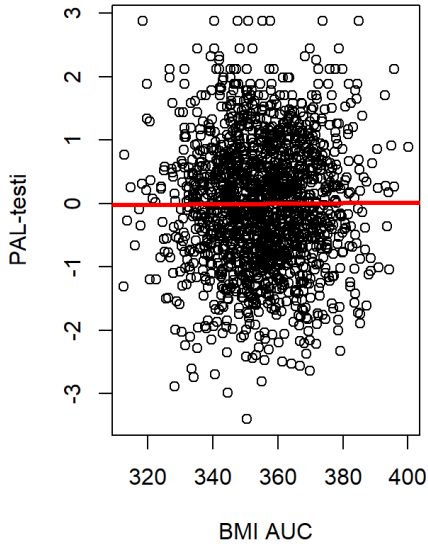
## Painoindeksi (BMI)

### PAL-testi

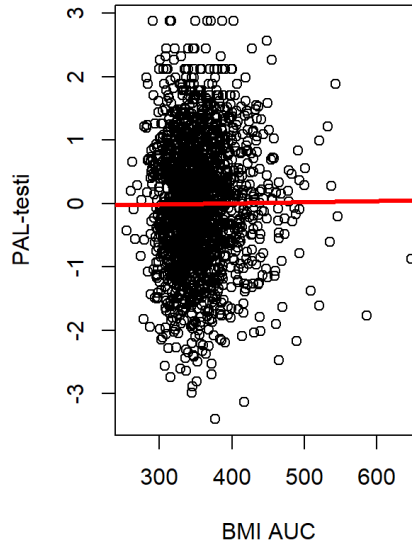
```
r_BMI1 <-
  twoSLS(xvar = analysisfull$bmi_6_24_auc, xname = "BMI AUC", xpred = "BMI_pred",
        zvar = analysisfull$BMI_UKBB_prscs_SD, zname = "Standardized BMI PRSCS",
        yvar = analysisfull$palfac11, yname = "PAL-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3939  -0.6763  -0.0013   0.6713   2.8950
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.1660567  0.6305414  -0.263   0.792
## xpred         0.0004592  0.0017733   0.259   0.796
##
## (Dispersion parameter for gaussian family taken to be 1.001882)
##
## Null deviance: 1656.2  on 1654  degrees of freedom
## Residual deviance: 1656.1  on 1653  degrees of freedom
## (371 observations deleted due to missingness)
## AIC: 4703.8
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept) -1.401895197  1.069781829
## xpred        -0.003016289  0.003934785
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4030  -0.6720  -0.0002   0.6752   2.8867
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0644728  0.2069252  -0.312   0.755
## xvar         0.0001814  0.0005785   0.314   0.754
##
## (Dispersion parameter for gaussian family taken to be 0.9982231)
##
## Null deviance: 1842.8  on 1847  degrees of freedom
## Residual deviance: 1842.7  on 1846  degrees of freedom
## (178 observations deleted due to missingness)
## AIC: 5245.1
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept) -0.4700387109  0.341093188
## xvar        -0.0009523959  0.001315165
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_BMI <- tsls(palfac11~bmi_6_24_auc, ~BMI_UKBB_prscs_SD, data = analysisfull)
summary(tsls_BMI)
```

```
##
## 2SLS Estimates
##
## Model Formula: palfac11 ~ bmi_6_24_auc
##
## Instruments: ~BMI_UKBB_prscs_SD
##
## Residuals:
##   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.406290 -0.676240 -0.001198  0.000000  0.672350  2.907866
##
##           Estimate   Std. Error  t value Pr(>|t|)
## (Intercept) -0.1686447097  0.6406195441 -0.26325  0.79239
## bmi_6_24_auc  0.0004663681  0.0018010183  0.25895  0.79571
##
## Residual standard error: 1.0010851 on 1653 degrees of freedom
```

```
confint(tsls_BMI)
```

```
##           2.5 %   97.5 %
## (Intercept) -1.424235944  1.086946525
## bmi_6_24_auc -0.003063563  0.003996299
```

```
r_BMI2 <-
  inverse_variance_weighted(analysisfull$bmi_6_24_auc, analysisfull$BMI_UKBB_prscs_SD,
    analysisfull$palfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI p-value
## IVW 0.000 0.002 -0.003, 0.004 0.796
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 241.6.
## [1] 0.000459248
```

```
RESULTS_BMI <- rows_patch(r_BMI1, r_BMI2, by="Test")
RESULTS_BMI[3,2] <- tsls_BMI$coefficients[2]
RESULTS_BMI
```

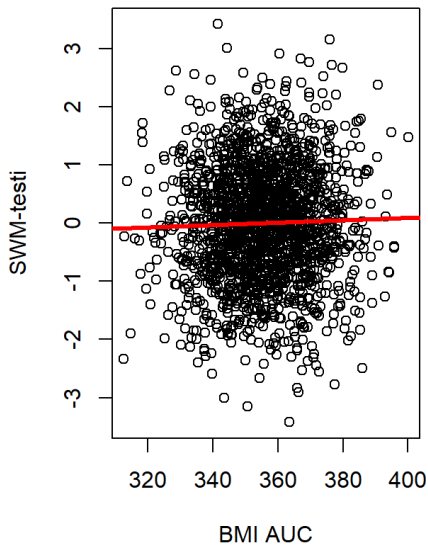
```
##          Test          Effect
## 1 Basic LM 0.0001813847
## 2 MR 2SLS1 0.0004592480
## 3 MR 2SLS2 0.0004663681
## 4 MR IVW 0.0004592480
## 5 MR TRIANGLE 0.0004592480
```

## SWM-Testi

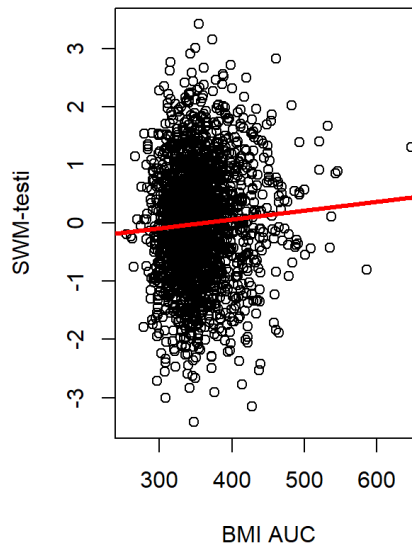
```
r_BMI1_swm <-
  twoSLS(xvar = analysisfull$bmi_6_24_auc, xname = "BMI AUC", xpred = "BMI_pred",
        zvar = analysisfull$BMI_UKBB_prscs_SD, zname = "Standardized BMI PRSCS",
        yvar = analysisfull$swmfac11, yname = "SWM-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4390  -0.6708  -0.0093   0.6824   3.4498
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.718794   0.605716  -1.187   0.236
## xpred        0.002023   0.001702   1.189   0.235
##
## (Dispersion parameter for gaussian family taken to be 0.9953427)
##
## Null deviance: 1792.0  on 1800  degrees of freedom
## Residual deviance: 1790.6  on 1799  degrees of freedom
## (225 observations deleted due to missingness)
## AIC: 5106.6
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -1.905975393  0.46838722
## xpred        -0.001313371  0.00536023
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4108  -0.6729   0.0056   0.6811   3.4235
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.5423830   0.1987243  -2.729  0.00640 **
## xvar         0.0015261   0.0005556   2.747  0.00608 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9949921)
##
## Null deviance: 2006.4  on 2010  degrees of freedom
## Residual deviance: 1998.9  on 2009  degrees of freedom
## (15 observations deleted due to missingness)
## AIC: 5700.9
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.9318754891 -0.152890412
## xvar         0.0004370681  0.002615093
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_BMI_swm <- tsls(swmfac11~bmi_6_24_auc, ~BMI_UKBB_prscs_SD, data = analysisfull)
summary(tsls_BMI_swm)
```

```
##
## 2SLS Estimates
##
## Model Formula: swmfac11 ~ bmi_6_24_auc
##
## Instruments: ~BMI_UKBB_prscs_SD
##
## Residuals:
##   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.40771 -0.67119  0.01456  0.00000  0.68839  3.42352
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.712524961  0.600588830 -1.18638  0.23563
## bmi_6_24_auc  0.002005902  0.001688137  1.18823  0.23490
##
## Residual standard error: 0.9979074 on 1799 degrees of freedom
```

```
confint(tsls_BMI_swm, level=0.95)
```

```
##           2.5 %    97.5 %
## (Intercept) -1.889657438  0.464607515
## bmi_6_24_auc -0.001302786  0.005314591
```

```
r_BMI2_swm <-
  inverse_variance_weighted(analysisfull$bmi_6_24_auc, analysisfull$BMI_UKBB_prscs_SD,
    analysisfull$swmfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI p-value
## IVW 0.002 0.002 -0.001, 0.005 0.235
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 241.6.
## [1] 0.00202343
```

```
RESULTS_BMI <- rows_patch(r_BMI1_swm, r_BMI2_swm, by="Test")
RESULTS_BMI[3,2] <- tsls_BMI_swm$coefficients[2]
RESULTS_BMI
```

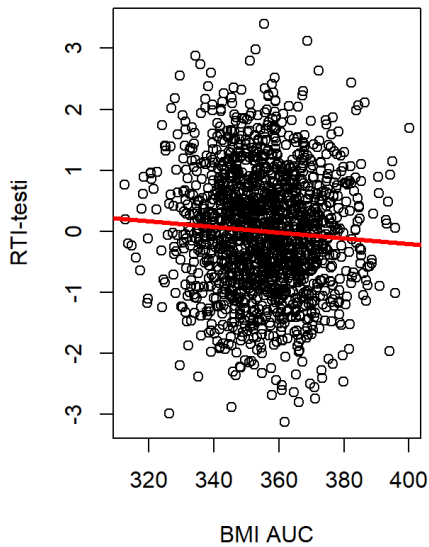
```
##          Test      Effect
## 1  Basic LM 0.001526080
## 2  MR 2SLS1 0.002023430
## 3  MR 2SLS2 0.002005902
## 4  MR IVW 0.002023430
## 5 MR TRIANGLE 0.002023430
```

## RTI-Testi

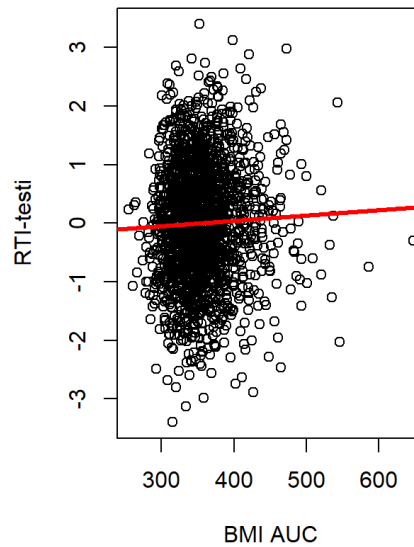
```
r_BMI1_rti <-
  twoSLS(xvar = analysisfull$bmi_6_24_auc, xname = "BMI AUC", xpred = "BMI_pred",
        zvar = analysisfull$BMI_UKBB_prscs_SD, zname = "Standardized BMI PRSCS",
        yvar = analysisfull$rtifac11, yname = "RTI-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1214  -0.6721  -0.0097   0.6825   3.3885
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.648680   0.635464   2.594  0.00956 **
## xpred        -0.004621   0.001787  -2.586  0.00980 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9966669)
##
##    Null deviance: 1629.2  on 1629  degrees of freedom
## Residual deviance: 1622.6  on 1628  degrees of freedom
## (396 observations deleted due to missingness)
## AIC: 4624.3
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept)  0.403193542  2.894166578
## xpred        -0.008123228 -0.001118352
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3588  -0.6823   0.0069   0.6685   3.3981
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.3245367   0.2080644  -1.56   0.119
## xvar         0.0009131   0.0005817   1.57   0.117
##
## (Dispersion parameter for gaussian family taken to be 0.9972719)
##
##    Null deviance: 1817.5  on 1821  degrees of freedom
## Residual deviance: 1815.0  on 1820  degrees of freedom
## (204 observations deleted due to missingness)
## AIC: 5169.6
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %      97.5 %
## (Intercept) -0.732335570  0.083262083
## xvar        -0.000226989  0.002053217
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_BMI_rti <- tsls(rtifac11~bmi_6_24_auc, ~BMI_UKBB_prscs_SD, data = analysisfull)
summary(tsls_BMI_rti)
```

```
##
## 2SLS Estimates
##
## Model Formula: rtifac11 ~ bmi_6_24_auc
##
## Instruments: ~BMI_UKBB_prscs_SD
##
## Residuals:
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.232994 -0.690375 -0.007151  0.000000  0.685560  3.522735
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.675356035  0.663120551  2.52647 0.011616 *
## bmi_6_24_auc -0.004694904  0.001864428 -2.51815 0.011893 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.025151 on 1628 degrees of freedom
```

```
confint(tsls_BMI_rti, level=0.95)
```

```
##           2.5 %      97.5 %
## (Intercept)  0.375663638  2.975048432
## bmi_6_24_auc -0.008349117 -0.001040692
```

```
r_BMI2_rti <-
  inverse_variance_weighted(analysisfull$bmi_6_24_auc, analysisfull$BMI_UKBB_prscs_SD,
    analysisfull$rtifac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
## IVW   -0.005    0.002 -0.008, -0.001  0.010
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 241.6.
## [1] -0.00462079
```

```
RESULTS_BMI <- rows_patch(r_BMI1_rti, r_BMI2_rti, by="Test")
RESULTS_BMI[3,2] <- tsls_BMI_rti$coefficients[2]
RESULTS_BMI
```

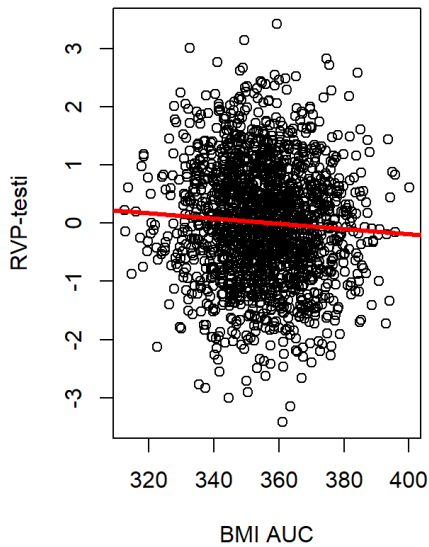
```
##           Test           Effect
## 1 Basic LM  0.0009131142
## 2 MR 2SLS1 -0.0046207901
## 3 MR 2SLS2 -0.0046949043
## 4 MR IVW  -0.0046207901
## 5 MR TRIANGLE -0.0046207901
```

## RVP-Testi

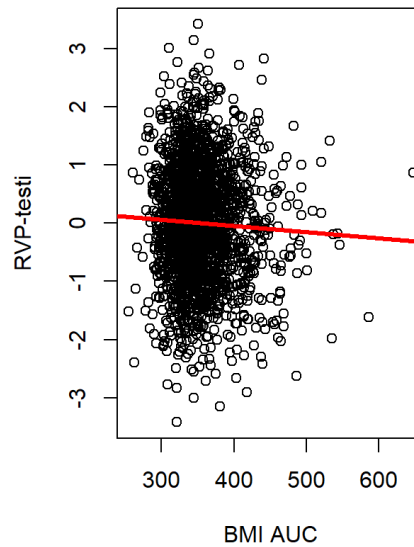
```
r_BMI1_rvp <-
  twoSLS(xvar = analysisfull$bmi_6_24_auc, xname = "BMI AUC", xpred = "BMI_pred",
        zvar = analysisfull$BMI_UKBB_prsccs_SD, zname = "Standardized BMI PRSCS",
        yvar = analysisfull$rvpfac11, yname = "RVP-testi", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4048  -0.6636   0.0076   0.6650   3.4221
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.627438   0.610274   2.667  0.00773 **
## xpred        -0.004541   0.001715  -2.647  0.00818 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9941602)
##
## Null deviance: 1763.6 on 1768 degrees of freedom
## Residual deviance: 1756.7 on 1767 degrees of freedom
## (257 observations deleted due to missingness)
## AIC: 5013.8
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept)  0.431322963  2.823553257
## xpred        -0.007902942 -0.001179127
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4533  -0.6717   0.0012   0.6752   3.4118
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.3778928   0.2008251   1.882  0.0600 .
## xvar        -0.0010637   0.0005618  -1.894  0.0584 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9968977)
##
## Null deviance: 1970.5 on 1974 degrees of freedom
## Residual deviance: 1966.9 on 1973 degrees of freedom
## (51 observations deleted due to missingness)
## AIC: 5602.7
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept) -0.015717095  7.715028e-01
## xvar        -0.002164756  3.728695e-05
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_BMI_rvp <- tsls(rvpfac11~bmi_6_24_auc, ~BMI_UKBB_prscs_SD, data = analysisfull)
summary(tsls_BMI_rvp)
```

```
##
## 2SLS Estimates
##
## Model Formula: rvpfac11 ~ bmi_6_24_auc
##
## Instruments: ~BMI_UKBB_prscs_SD
##
## Residuals:
##   Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -3.58456 -0.68620 -0.01122  0.00000  0.67327  3.38074
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.619405928  0.612936805  2.64204 0.0083135 **
## bmi_6_24_auc -0.004519459  0.001723149 -2.62279 0.0087963 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.006426 on 1767 degrees of freedom
```

```
confint(tsls_BMI_rvp, level=0.95)
```

```
##           2.5 %      97.5 %
## (Intercept)  0.41807187  2.820739990
## bmi_6_24_auc -0.00789677 -0.001142148
```

```
r_BMI2_rvp <-
  inverse_variance_weighted(analysisfull$bmi_6_24_auc, analysisfull$BMI_UKBB_prscs_SD,
    analysisfull$rvpfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
##   IVW   -0.005    0.002 -0.008, -0.001  0.008
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 241.6.
## [1] -0.004541035
```

```
RESULTS_BMI <- rows_patch(r_BMI1_rvp, r_BMI2_rvp, by="Test")
RESULTS_BMI[3,2] <- tsls_BMI_rvp$coefficients[2]
RESULTS_BMI
```

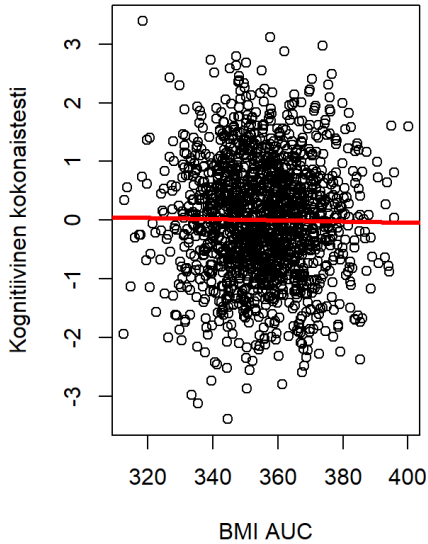
```
##           Test           Effect
## 1 Basic LM -0.001063734
## 2 MR 2SLS1 -0.004541035
## 3 MR 2SLS2 -0.004519459
## 4 MR IVW -0.004541035
## 5 MR TRIANGLE -0.004541035
```

## Kognitiivinen kokonaissumma (cognitive score)

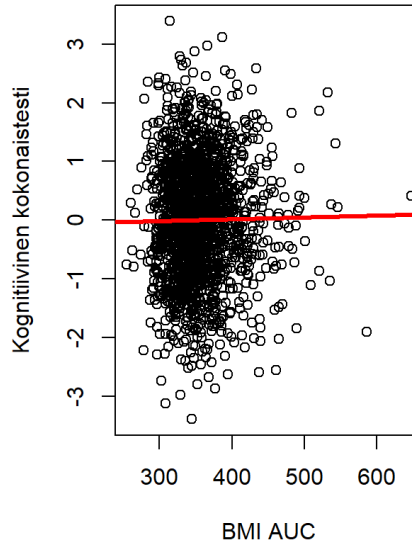
```
r_BMI1_cog <-
  twoSLS(xvar = analysisfull$bmi_6_24_auc, xname = "BMI AUC", xpred = "BMI_pred",
        zvar = analysisfull$BMI_UKBB_prscs_SD, zname = "Standardized BMI PRSCS",
        yvar = analysisfull$cogfac11, yname = "Kognitiivinen kokonaistesti", d=analysisfull)
```

```
##
## Call:
## glm(formula = yvar ~ xpred, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4082  -0.6708   0.0051   0.6809   3.3479
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.369569   0.640223   0.577   0.564
## xpred        -0.001024   0.001800  -0.569   0.570
##
## (Dispersion parameter for gaussian family taken to be 1.001768)
##
## Null deviance: 1611.2 on 1609 degrees of freedom
## Residual deviance: 1610.8 on 1608 degrees of freedom
## (416 observations deleted due to missingness)
## AIC: 4575.8
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept) -0.885245483  1.624384421
## xpred        -0.004552914  0.002504481
##
## Call:
## glm(formula = yvar ~ xvar, data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3882  -0.6695   0.0010   0.6758   3.4043
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.1104335   0.2099201  -0.526   0.599
## xvar         0.0003109   0.0005872   0.529   0.597
##
## (Dispersion parameter for gaussian family taken to be 0.9984514)
##
## Null deviance: 1793.5 on 1797 degrees of freedom
## Residual deviance: 1793.2 on 1796 degrees of freedom
## (228 observations deleted due to missingness)
## AIC: 5103.7
##
## Number of Fisher Scoring iterations: 2
##
##              2.5 %       97.5 %
## (Intercept) -0.5218692756  0.301002350
## xvar        -0.0008400447  0.001461818
```

## Altisteen kausaalinen vaikutus



## Alkuperäinen altisteen yhteys



```
tsls_BMI_cog <- tsls(cogfac11~bmi_6_24_auc, ~BMI_UKBB_prscs_SD, data = analysisfull)
summary(tsls_BMI_cog)
```

```
##
## 2SLS Estimates
##
## Model Formula: cogfac11 ~ bmi_6_24_auc
##
## Instruments: ~BMI_UKBB_prscs_SD
##
## Residuals:
##   Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -3.40835 -0.67590  0.01024  0.00000  0.68006  3.34295
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.374875641  0.650071586  0.57667  0.56424
## bmi_6_24_auc -0.001039112  0.001828056 -0.56842  0.56983
##
## Residual standard error: 1.0016967 on 1608 degrees of freedom
```

```
confint(tsls_BMI_cog, level=0.95)
```

```
##           2.5 %      97.5 %
## (Intercept) -0.899241254  1.648992536
## bmi_6_24_auc -0.004622036  0.002543812
```

```
r_BMI2_cog <-
  inverse_variance_weighted(analysisfull$bmi_6_24_auc, analysisfull$BMI_UKBB_prscs_SD,
    analysisfull$cogfac11)
```

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 1
##
## -----
## Method Estimate Std Error 95% CI      p-value
##   IVW   -0.001    0.002 -0.005, 0.003  0.569
## -----
## Residual standard error = 1.000
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Heterogeneity is not calculated when weights are penalized, or when there is only one variant in the analysis.
## F statistic = 241.6.
## [1] -0.001024217
```

```
RESULTS_BMI <- rows_patch(r_BMI1_cog, r_BMI2_cog, by="Test")
RESULTS_BMI[3,2] <- tsls_BMI_cog$coefficients[2]
RESULTS_BMI
```

##	Test	Effect
## 1	Basic LM	0.0003108864
## 2	MR 2SLS1	-0.0010242167
## 3	MR 2SLS2	-0.0010391118
## 4	MR IVW	-0.0010242167
## 5	MR TRIANGLE	-0.0010242167