

Meta-analyysin ja metaregression soveltaminen fyysistä aktiivisuutta mittaaviin tutkimuksiin

Tilastotieteen pro gradu -tutkielma

21.12.2017

Jaakko Immonen

Matematiikan ja tilastotieteen laitos

Jyväskylän yliopisto

Tiivistelmä: Jaakko Immonen, *Meta-analyysin ja metaregression soveltaminen fyysistä aktiivisuutta mittaaviin tutkimuksiin*, Tilastotieteen pro gradu -tutkielma, 38 s. + liitteitä 4 s., Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, 21.12.2017.

Tieteellistä tietoa tuotetaan jatkuvasti lisää, mutta sen yhdistäminen aiempaan tietämykseen ei ole ongelmatonta. Tilastotieteellisesti tätä pulmaa lähestytään meta-analyysin ja metaregression avulla. Meta-analyysi on systemaattisessa kirjallisuuskatsauksessa kerätyn aineiston tilastollista yhteenvedoa; useista estimaateista luottamusväleinen muodostetaan yksi keskimääräinen estimaatti ja sen luottamusväli. Huomattavaa on, että meta-analyysin aineisto on eräänlainen aineistojen aineisto; havaintoyksiköitä eivät ole tutkimuksiin osallistuvat koehenkilöt vaan itse tutkimukset. Metaregressio on meta-analyysiä, jossa yhteenvedo tehdään hyödyntämällä tutkimusten taustamuuttujia eli kovariaatteja. Metaregressio on lineaaristen (sekä)mallien erikoistapaus.

Tutkielmassa meta-analyysin menetelmiä sovelletaan aineistoon, joka käsittelee vähän liikkuvien ihmisten fyysisen aktiivisuuden edistämistä erilaisin motivointikeinoin. Tutkimusten ominaispiirteiden, kuten valittujen motivointikeinojen, yhteyttä tuloksiin tutkitaan metaregressiolla. Oman haasteensa tuo aineistossa esiintyvä riippuvuus. Riippuvuus on seurausta yhteisistä kontrolliryhmistä, joihin eri motivointikeinoja käyttäviä ryhmiä verrataan; jokaisessa tutkimuksessa on täsmälleen yksi kontrolliryhmä mutta mahdollisesti useita koeryhmiä.

Tulosten perusteella motivointikeinot todella lisäävät vähän liikkuvien ihmisten fyysistä aktiivisuutta jopa siten, että vaikutus jossain määrin säilyy intervention päätyttyä. Eri motivointikeinojen kesken ei näytä olevan huomattavia eroja. Viitteitä muiden taustamuuttujien yhteydestä tuloksiin on, mutta niihin on suhtauduttava varauksella; monta testattavaa hypoteesia voi johtaa merkitseviin tuloksiin sattumaltakin. Lisäksi metaregression tulokset vaihtelevat valittujen menetelmien mukana.

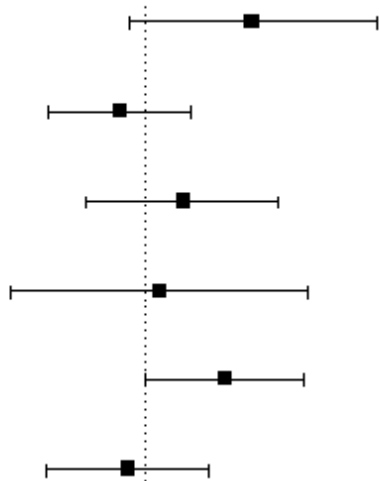
Sisältö

1	Johdanto	1
2	Sovellusaineisto	2
3	Meta-analyysi ja metaregressio	5
3.1	Tilastollisesti riippumattomat vaikutukset	5
3.1.1	Kaksitasoinen sekamalli	5
3.1.2	Kolmitasoinen sekamalli	7
3.2	Tilastollisesti riippuvat vaikutukset	8
3.3	Sekamallin ja kiinteiden vaikutusten mallin ero	9
3.4	Keskiarvojen vertailuun sopivia vaikutuksia	9
3.4.1	Tilastollisesti riippumattomat vaikutukset	10
3.4.2	Yhteisestä kontrolliryhmästä aiheutuva riippuvuus	12
3.5	Graafiset tarkastelut ja julkaisuharha	12
4	Sovellusaineiston meta-analyysin toteutus	15
5	Meta-analyysin tulokset sovellusaineistossa	16
5.1	Loppuvaihe	16
5.2	Seurantavaihe	18
5.3	Poikkeavat havainnot	19
5.4	Julkaisuharhan arviointi ja huomioiminen	20
5.5	Riippumattomat vaikutukset	24
5.6	Standardoitu keskiarvojen erotus	25
6	Pohdintaa	29
	Lähteet	32
	Liitteet	39

1 Johdanto

Tieteellisen tiedon kertyminen lienee ensisijaisesti myönteinen asia. Kertyvällä tiedolla voidaan mullistaa käsityksiämme maailmasta ja sen ilmiöistä. Esimerkiksi sopii vauvojen hoitosuositusten muuttuminen 1990-luvulla (ks. Gilbert, Salanti, Harden & See 2005). Tuohon asti monet lastenlääkärit suosittivat vanhempia laittamaan vauvansa nukkumaan vatsalleen. Sitten suositus on muuttunut, sillä vatsallaan nukkumisen on todettu johtaneen kymmenien tuhansien vauvojen kätkytkuolemaan (Gilbert ym. 2005). Nykyään pidetään järkevänä laittaa vauva nukkumaan selälleen.

Kertyvän tiedon käsittely ei ole kuitenkaan ongelmaton. Voi olla, että samaan tutkimuskysymykseen liittyvät tutkimustulokset ovat johdonmukaisia, mutta yhdenkään yksittäisen tutkimuksen näyttö ei riitä halutun väitteen todentamiseen. Toisaalta eri koeasetelmien ja sattuman vuoksi tutkimusten tulokset ja johtopäätökset saattavat erota toisistaan huomattavasti. Siitä huolimatta on luultavaa, että taustalla on jonkinlainen riippumaton totuus, jota kaikilla tutkimuksilla tavoitellaan.



Kuva 1: Kuudesta eri tutkimuksesta saadut estimaatit ja luottamusvälit. Aineisto on simuloitu.

Analyysin kohteena on eräänlainen aineistojen aineisto; havaintoyksiköitä ovat tutkimukset itsessään toisin kuin yksittäisissä tutkimuksissa, joissa havaintoyksiköitä ovat esimerkiksi nukkuvat vauvat tai lääketestaukseen osallistuvat koehenkilöt.

Tässä työssä esitellään ensin Jyväskylän yliopiston terveystieteiden laitokselta saatu sovellusaineisto, joka käsittelee erilaisten motivointikeinojen merkitystä fyysisen

Systemaattinen kirjallisuuskatsaus pureutuu kertyvän tiedon pulmaan. Aluksi systemaattisessa kirjallisuuskatsauksessa etsitään tiettyjä hakukriteerejä käyttäen joukko tutkimuksia. Sitten päätetään, mitkä tutkimuksista sisällytetään varsinaiseen analyysiin. Lopuksi irrallisista tiedonpalasista tehdään yhteenveto, joka ottaa huomioon tutkimuskohtaiset tulokset sekä niiden epävarmuuden ja kenties kunkin tutkimuksen ominaispiirteitä. Yhteenvedossa voidaan hyödyntää *meta-analyysiä*, jolla tarkoitetaan kirjallisuuskatsauksen aiemmissa vaiheissa kerätyn aineiston tilastollista analyysiä. (Borenstein, Hedges, Higgins & Rothstein 2009.) Meta-analyysissä on pelkistetysti kyse yhden keskimääräisen estimaatin ja sen luottamusvälin johtamisesta useista eri tutkimuksista saaduista estimaateista ja luottamusväleistä (ks. kuva 1). Meta-regressio on meta-analyysiä, jossa tutkimusten taustamuuttujia eli kovariaatteja on mukana mallissa. Kyse on lineaarisen (seka)mallin erikoistapauksesta.

aktiivisuuden lisäämisessä vähän liikkuvien aikuisten kuntoutuksessa. Fyysinen aktiivisuus ennaltaehkäisee muun muassa sydän- ja verisuonisairauksien syntymistä ja diabetekseen sairastumista (Bauman 2004). Siten ei ole yhdentekevää, liikkuvatko ihmiset vai eivät. Tarkoitus on selvittää, kuinka ihmisiä voidaan motivoida liikkumaan ja ovatko jotkin motivointikeinot muita vaikuttavampia. Motivointikeinojen ohella monien muidenkin kovariaattien, kuten intervention keston, yhteyttä tuloksiin tarkastellaan. Kiinnostuksen kohteena on myös, kuinka hyvin vaikutus säilyy varsinaisen intervention jälkeen. Interventiolla tarkoitetaan tutkimukseen osallistuviin henkilöihin kohdistettua toimintaa, joilla heidän fyysistä aktiivisuuttaan pyritään lisäämään.

Aineiston kuvailun jälkeen käydään läpi sen analysoinnin kannalta olennaista meta-analyysin ja metaregression teoriaa. Lukuisien kovariaattien lisäksi aineistolle ovat ominaisia tutkimukset, joissa monta koeryhmää verrataan yhteen kontrolliryhmään. Riippuvuus on otettava huomioon meta-analyysimalleissa. Toisaalta on kiinnostavaa kokeilla, kuinka vahvasti riippuvuus vaikuttaa tuloksiin, joten vertailun vuoksi analyysit tehdään myös riippumattomuusoletusta käyttäen. Analyyseissa otetaan huomioon poikkeavat havainnot ja valikoivan raportoinnin mahdollisesti tuottama julkaisuharha. Lopuksi esitellään analyysien tulokset.

2 Sovellusaineisto

Tutkielmassa käytettävä aineisto on saatu Jyväskylän yliopiston terveystieteiden laitokselta. Tavoitteena on selvittää, kuinka ihmisiä voidaan motivoida liikkumaan; onko esimerkiksi tavoitteiden asettamisesta, palautteesta, sosiaalisesta tuesta, palkkioista tai rangaistuksista motivointikeinoina hyötyä? Useisiin eri tietokantoihin kohdistetulla haulla löytyi 42 tutkimusta, jotka on arvioitu tutkimuskysymykseen ja meta-analyysiin sopiviksi. Kaikissa näissä tutkimuksissa on tehty satunnaistettu vertailukoe; motivointikeinojen merkitystä on tutkittu jakamalla osallistujat satunnaisesti kontrolliryhmään ja yhteen tai useampaan koeryhmään. Kontrolliryhmän osallistujat ovat saaneet tavanomaisen intervention ilman motivointikeinoja. Koeryhmän interventio on muuten samanlainen kuin kontrolliryhmän interventio, mutta koeryhmän interventioon kuuluu ainakin yksi motivointikeino.

Aineisto sisältää 28 kovariaattia, jotka kaikki luetellaan taulukossa 1. Nimenomaan eri motivointikeinot, kuten tavoitteet ja suunnittelu sekä sosiaalinen tuki, ovat keskeisimpiä kovariaatteja. Motivointikeinoja on yhteensä 17, ja niistä kuutta yleisintä hyödynnetään sellaisenaan dummy-muuttujina. Loput 11 keinoa yhdistetään yhdeksi dummy-muuttujaksi; joko ainakin yhtä niistä on käytetty tai sitten ei. Useimmat motivointikeinot, joskaan eivät kaikki, jakautuvat alakohtiin. Esimerkiksi tavoitteet ja suunnittelu jakautuu yhdeksään eri alakohtaan, joihin lukeutuvat muun muassa ongelmien ratkaisu, toiminnan suunnittelu ja sitoutuminen. Laskemalla alakohtien dummy-muuttujat yhteen saadaan motivointikeinojen lukumäärä, joka on yksi kovariaatti. Kyse on siis alakohtien lukumäärästä, ei 17 ylätasen motivointikeinojen

lukumäärästä.

Taulukko 1: Lista kovariaateista selityksineen.

Kovariaatti	Selitys
goals	Onko <i>tavoitteet ja suunnittelu</i> motivointikeinona?
feedback	Onko <i>palaute ja tarkkailu</i> motivointikeinona?
support	Onko <i>sosiaalinen tuki</i> motivointikeinona?
shaping	Onko <i>tiedon muokkaus</i> motivointikeinona?
repetition	Onko <i>toisto ja korvaaminen</i> motivointikeinona?
reward	Onko <i>palkkio ja rangaistus</i> motivointikeinona?
other	Onko jokin muu motivointikeino käytössä?
nbct	Alatason motivointikeinojen lukumäärä
multibct	Onko käytössä 1–5 motivointikeinojen alakohtaa vai enemmän?
post	Onko ainakin osa interventiosta annettu postitse?
face	Onko ainakin osa interventiosta annettu kasvotusten?
phone	Onko ainakin osa interventiosta annettu puhelimitse?
web	Onko ainakin osa interventiosta annettu internetin välityksellä?
multidel	Onko intervention antamisessa hyödynnetty yhtä vai monta tapaa?
multispec	Onko intervention antanut yksi vai useampi asiantuntija?
delto	Onko interventio annettu yksilöille, ryhmille vai molemmille?
paonly	Onko intervention ainoa tavoite aktiivisuuden lisääminen?
design	Mitä vaiheita interventiossa on ja milloin (neljä luokkaa)?
objective	Onko aktiivisuutta mitattu objektiivisesti?
health	Ovatko osallistujat terveitä, sairaita vai sairaudesta selvinneitä?
control	Kuinka vahvan ohjeistuksen kontrolliryhmä on saanut (viisi luokkaa)?
quality	Tutkimuksen toteutuksen ja raportoinnin laatu asteikolla 0–12
theory	Millaisia teorioita koeryhmän intervention taustalla on (neljä luokkaa)?
intensity	Koeryhmän kontaktiaika minuutteina intervention aikana
durint	Koeryhmän kontaktien lukumäärä tutkimuksen aikana
lengthint	Interventiovaiheen kesto viikkoina
durfollow	Koeryhmän kontaktien lukumäärä seurantavaiheessa
lengthfollow	Seurantavaiheen kesto viikkoina

Eräs olennainen kovariaatti on kontrolliryhmän ohjeistuksen taso. Se on järjestysasteikollinen luokkineen 1 = ”ei ohjeistusta”, 2 = ”yleinen elämäntapaohjeistus”, 3 = ”ohjeistus fyysiseen aktiivisuuteen”, 4 = ”jonkin verran yhteisiä aktiviteetteja” ja 5 = ”paljon yhteisiä aktiviteetteja”. Tutkimuksen toteutuksen ja raportoinnin laatua on arvioitu 12 kriteerillä, joista kunkin tutkimus joko täyttää tai ei täytä. Kriteereissä käydään läpi esimerkiksi satunnaistusta, sokkoutusta ja valikoivaa raportointia koskevia kysymyksiä. Mitä luotettavampi tutkimus on, sitä laadukkaampi se on saaden enemmän pisteitä. Kriteerit itsessään ovat objektiivisia, mutta näkemys niiden täyttymisestä on enemmän tai vähemmän subjektiivinen. Kaksi asiantuntijaa ovat toisistaan riippumattomasti arvioineet tutkimukset. Laatua mittaava kovariaatti on näiden kahden arvion keskiarvo; sen mahdolliset arvot ovat 0–12 puolen

pisteen välein.

Useimmat koeryhmän interventiot perustuvat ainakin yhteen ihmisten käytöstä selittävään teoriaan, joista esimerkkinä mainittakoon sosiokognitiivinen teoria. Chandra ym. (2017) kuvailevat kaikkia tutkimuksista tunnistettuja teorioita. Osa teorioista on hyvin yleisiä, kun taas muut ovat suhteellisen harvinaisia. Näin saadaan neliluokkainen muuttuja; koeryhmän interventio voi perustua vain yleisiin teorioihin, vain harvinaisiin teorioihin, sekä yleisiin että harvinaisiin teorioihin tai sitten sillä ei ole minkäänlaista teoriapohjaa.

Tutkimukset voidaan jakaa toteutustapansa mukaan neljään luokkaan (ks. kuva 2). Kaikille tutkimuksille on yhteistä interventiovaihe, jossa koehenkilöihin ollaan aktiivisesti yhteydessä. Ensimmäisessä toteutustavassa interventiota seuraa loppumittaus ja seurantavaihe, joka päättyy seurantamittauksiin. Seurantavaiheessa koehenkilöihin ei pidetä yhteyttä. Toisessa tavassa toimitaan muuten kuin edellä, mutta loppumittauksia ei ole tehty tai niitä ei ole raportoitu. Kolmas tapa poikkeaa ensimmäisestä vain seurantavaiheeltaan; koehenkilöihin ollaan yhteydessä esimerkiksi puhelimitse tai sähköpostitse. Neljäs tapa vastaa muuten ensimmäistä, mutta seurantamittauksia ei ole tehty tai niitä ei saatavilla.

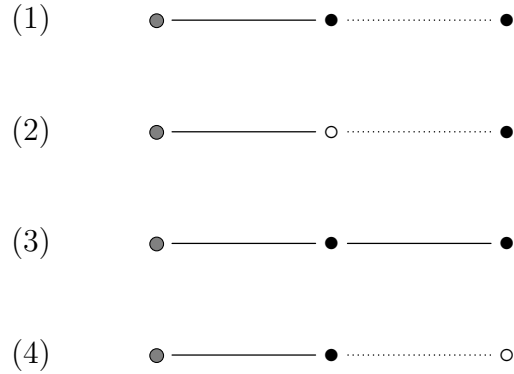
Loppu- ja seurantamittausten ohella monissa tutkimuksissa on tehty myös interventiota edeltävät alkumittaukset. Lukuisien syiden vuoksi alkumittauksia ei oteta analyysiin mukaan. Ensinnäkin koe- ja kontrolliryhmät on satunnaistettu ja otoskoko on tavallisesti useita kymmeniä osallistujia per ryhmä, joten ryhmien voidaan olettaa olevan alkutilanteessa likimain yhtä aktiivisia. Toisekseen tarkoitus on vertailla nimenomaan koe- ja kontrolliryhmän interventioiden tehoa toisiinsa eikä interventioiden tehoa suhteessa alkutilanteeseen. Voidaan olettaa, että jo kontrolliryhmän interventio lisää fyysistä aktiivisuutta tai ei ainakaan vähennä sitä. Jos koeryhmän interventio osoittautuu merkittävästi paremmaksi kuin kontrolliryhmän interventio, voidaan hyvällä syyllä uskoa sen myös lisäävän fyysistä aktiivisuutta alkutilanteeseen nähden.

Fyysistä aktiivisuutta on mitattu lukuisilla eri mittayksiköillä. Lähtökohtana on usein käytetty esimerkiksi askelten lukumäärää, aktiivisten minuuttien määrää tai kulutettuja kilokaloreita, jotka lopuksi suhteutetaan johonkin aikayksikköön, kuten päiviin tai viikkoihin. Varsinaisia mittayksiköitä ovat muun muassa aktiiviset minuutit yhtä viikkoa kohden ja askelten lukumäärä per päivä. Monessa aktiivisuuden yksikössä on mukana MET eli lepoaineenvaihdunnan kerrannainen. Yksi MET on

$$\frac{4.184 \text{ kJ}}{\text{kg} \cdot \text{t}},$$

ja se vastaa ihmisen energiankulutusta levossa painoon ja kuluneeseen aikaan suhteutettuna (Ainsworth ym. 2000, S498). Esimerkiksi nukkuminen vastaa 0.9 MET:ä, käveleminen noin viiden kilometrin tuntinopeudella 2.5 MET:ä ja juokseminen noin kahdentoista kilometrin tuntinopeudella 12.5 MET:ä (Ainsworth ym. 2000, S498, S511). MET-luvut voidaan kertoa kuluneella ajalla; jos osallistuja kävelee kolme-

Toteutustapa



Kuva 2: Kaavio tutkimusten eri toteutustavoista. Pisteet ovat ajanhetkiä ja niiden väri viittaa mittaamiseen; mittaus on suoritettu (musta) tai ei (valkoinen). Alkumittausten harmaat pisteet viittaavat siihen, että tutkimuksesta riippuen alkumittaukset on suoritettu tai jätetty suorittamatta. Yhtenäinen viiva kuvaa aktiivista ajanjaksoa ja katkoviiva ajanjaksoa, jolloin koehenkilöihin ei olla yhteydessä.

kymmentä minuuttia, se vastaa $30 \cdot 2.5 = 75$ MET-minuuttia tai $0.5 \cdot 2.5 = 1.25$ MET-tuntia. Esimerkiksi kuvissa 5 ja 6 ovat esillä kaikki käytetyt yksiköt.

3 Meta-analyysi ja metaregressio

Seuraavaksi esitellään Raudenbushia (2009) ja Viechtbaueria (2010) mukaillen meta-analyysissä käytettäviä malleja. Oletetaan että käytössä on n kappaletta tutkimuksia. Keskeisiä tunnuslukuja ovat *vaikutukset* $\theta_1, \dots, \theta_n$ ja niiden estimaatit y_1, \dots, y_n . Vaikutukset kuvaavat kahden ryhmän välistä eroa. Konkreettisesti vaikutus voi olla esimerkiksi viikossa käveltyjen askeleiden keskimääräinen erotus ryhmien välillä. Satunnaistuksen ansiosta ryhmät ovat muuten vertailukelpoiset keskenään, mutta niiden jäsenet saavat erilaisen intervention. Tämän vuoksi termin ”vaikutus”käyttö on usein aiheellista, vaikka kausaalitulkintojen tekemisessä tuleekin olla varovainen.

3.1 Tilastollisesti riippumattomat vaikutukset

3.1.1 Kaksitasoinen sekamalli

Olkoot y_1, \dots, y_n toisistaan riippumattomia vaikutuksen estimaatteja. Kiinteiden vaikutusten meta-analyysimalli on

$$y_i = \theta_i + \varepsilon_i, \quad (1)$$

missä θ_i on vaikutuksen estimaattia y_i vastaava todellinen vaikutus ja $\varepsilon_i \sim N(0, \sigma_i^2)$ on jäännöstermi. Jäännöstermien normaalisuudesta seuraa, että $y_i \sim N(\theta_i, \sigma_i^2)$. Varianssit $\sigma_1^2, \dots, \sigma_n^2$ ovat käytännössä tuntemattomia. Ne kuitenkin estimoidaan aineistosta, ja niitä käytetään kuin ne olisivat tunnettuja. Useimmiten meta-analyysissä hyödynnetään sekamallia, joka saadaan kaavasta (1), kun

$$\theta_i = \mu + u_i, \quad (2)$$

missä satunnaisvaikutus $u_i \sim N(0, \tau^2)$. Tämän seurauksena $y_i \sim N(\mu, \sigma_i^2 + \tau^2)$. Mallista estimoidaan keskimääräinen vaikutus μ ja *heterogeenisyys* τ^2 . Jos $\tau^2 = 0$, todelliset vaikutukset $\theta_1, \dots, \theta_n$ ovat homogeenisiä, eli μ on todellinen vaikutus jokaisessa tutkimuksessa.

Tyypillisesti tutkimuksilla on ominaispiirteitä, jotka erottavat ne toisistaan. Toisin sanoen tutkimukset ovat heterogeenisiä. Jos tiedossa on tutkimusten piirteitä kuvaavat muuttujat X_1, \dots, X_p , ne voidaan ottaa malliin mukaan kovariaateiksi. Tämä onnistuu niin sanotun *metaregressiomallin* avulla valitsemalla kaavassa (2)

$$\mu = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}. \quad (3)$$

Lopputulema on

$$y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + u_i + \varepsilon_i. \quad (4)$$

Tällöin τ^2 mittaa jäljelle jäävän *jäännösheterogeenisyyden* määrää, jota kovariaatit eivät selitä. Estimoitavia parametreja ovat jäännösheterogeenisyys τ^2 ja regressiokertoimet $\beta_0, \beta_1, \dots, \beta_p$.

Estimointi voidaan tehdä esimerkiksi suurimman uskottavuuden avulla (Raudenbush 2009, 308–310). Muotoillaan ensin malli (4) vektoreiden avulla. Toisin sanoen

$$y_i = \mathbf{X}_i^T \boldsymbol{\beta} + u_i + \varepsilon_i,$$

missä $\mathbf{X}_i = (1, X_{i1}, \dots, X_{ip})^T$ ja $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$. Suurimman uskottavuuden menetelmällä parametreille voidaan johtaa estimointiyhtälöt

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^n w_i \mathbf{X}_i \mathbf{X}_i^T \right)^{-1} \sum_{i=1}^n w_i \mathbf{X}_i y_i \quad (5)$$

ja

$$\hat{\tau}^2 = \frac{\sum_{i=1}^n w_i^2 \left[(y_i - \mathbf{X}_i^T \boldsymbol{\beta})^2 - \sigma_i^2 \right]}{\sum_{i=1}^n w_i^2}, \quad (6)$$

missä $w_i = (\sigma_i^2 + \tau^2)^{-1}$. Parametrivektori $\hat{\boldsymbol{\beta}}$ voidaan mieltää myös painotetuksi pienimmän neliösumman ratkaisuksi, jossa painot ovat kääntäen verrannollisia varianssiin nähden (Viechtbauer 2009, 5). Yhtälöitä ei voida ratkaista suljetussa muodossa, sillä regressiokertoimien estimaattorit riippuvat heterogeenisyydestä ja heterogeenisyyden estimaattori regressiokertoimista. Estimointi onnistuu kuitenkin esimerkiksi

asettamalla aluksi $\tau^2 = 0$, jonka avulla saadaan ensin laskettua $\hat{\beta}$ ja edelleen regressiokertoimien estimaattien avulla $\hat{\tau}^2$. Näin on saatu uusi arvio heterogeenisyydelle, minkä jälkeen voidaan yhtälöille (5) ja (6) saadaan jälleen uudet ratkaisut. Estimaattit saadaan toistamalla tätä menettelytapaa riittävän monta kertaa.

Piste-estimoinnin varmuutta voidaan tarkastella normaalisuuteen perustuvilla testeillä ja luottamusväleillä. Olettamalla heterogeenisuus tunnetuksi parametrien $\hat{\beta}$ kovarianssimatriisi on (Raudenbush 2009, 312)

$$\text{Cov}(\hat{\beta} \mid \hat{\tau}^2 = \tau^2) = \left(\sum_{i=1}^n w_i \mathbf{X}_i \mathbf{X}_i^T \right)^{-1}, \quad (7)$$

ja riittävän suurella otoskoollla n

$$\widehat{\text{Cov}}(\hat{\beta}) = \text{Cov}(\hat{\beta} \mid \hat{\tau}^2 = \tau^2). \quad (8)$$

Kunkin parametrin β_i , $i = 1, \dots, n$, keskivirhe on kovarianssimatriisin lävistäjän i :nnen alkion neliöjuuri. Heterogeenisyyden merkitsevyyden testaukseen on olemassa Cochranin Q-testi (Hedges & Olkin 1985).

3.1.2 Kolmitasoinen sekamalli

Kaksitasoinen sekamalli olettaa heterogeenisyyden olevan yksitasoista. Meta-analyysille on kuitenkin tyypillistä aineiston hierarkkinen rakenne; osa vaikutuksista voi olla peräisin esimerkiksi samoilta tutkijoilta tai samasta laboratorion (Konstantopoulos 2011, 61). Tämän tutkielman aineistossa hierarkkinen rakenne syntyy vaikutuksista, jotka ovat peräisin samasta tutkimuksesta. Esimerkiksi Demark-Vahnefriedilta ym. (2014) on loppuvaiheessa neljä vaikutusta.

Hierarkian vuoksi on aiheellista laajentaa kappaleen 3.1.1 malli kolmitasoiseksi sekamalliksi (Konstantopoulos 2011, 65). Kutsutaan hierarkian ylempiä tasoja (soveluksen tapauksessa tutkimuksia) *blokeiksi*, ja olkoon niitä $m < n$ kappaletta. Kolmitasoinen sekamalli ilman kovariaatteja i :nnelle vaikutukselle (joka kuuluu j :nteen blokkiin) on

$$y_{ij} = \mu + v_{0j} + u_{ij} + \varepsilon_{ij}, \quad (9)$$

missä $v_{0j} \sim N(0, \omega^2)$, $i = 1 \dots n$ ja $j = 1 \dots m$ on. Keskimääräinen vaikutus μ , satunnaisvaikutus $u_{ij} \sim N(0, \tau^2)$ ja jäännöstermi $\varepsilon_{ij} \sim N(0, \sigma_{ij}^2)$ vastaavat kaksitasoisen sekamallin parametreja μ , u_i ja ε_i . Satunnaisvaikutus v_{0j} on sama kaikille blokin j vaikutuksille, mutta satunnaisvaikutukset v_{0j} ja $v_{0j'}$ ovat mahdollisesti erisuuret blokeille j ja j' . Kappaleessa 3.1.1 on käytetty ilmausta ”tutkimusten heterogeenisuus”, mutta kolmitasoisen sekamallin yhteydessä on järkevämpää puhua *vaikutusten* heterogeenisyydestä, joka on kahden heterogeenisyydskomponentin summa $\tau^2 + \omega^2$.

Kolmitasoinen sekamalli laajenee regressiomalliksi samaan tapaan kuin kaksitasoinen sekamalli (Konstantopoulos 2011, 65):

$$y_{ij} = \beta_{00} + \beta_{1j}X_{ij1} + \dots + \beta_{pj}X_{ijp} + v_{0j} + u_{ij} + \varepsilon_{ij}, \quad (10)$$

missä X_{ij1}, \dots, X_{ijp} ovat kovariaatteja ja $\beta_{00}, \beta_1 \dots \beta_p$ regressiokertoimia. Konstantopoulos (2011, 65) käyttää myös blokkikohtaisia kovariaatteja, jolloin samaan blokkiin kuuluviin vaikutuksiin liitetään sama kovariaatin arvo. Tämä on hyödyllistä, jos halutaan arvioida erikseen blokkikohtaisia keskimääräisiä vaikutuksia. Tässä tutkielmassa kiinnostuksen kohteena on kuitenkin vain kaikista vaikutuksista estimoitava keskimääräinen vaikutus, joten kovariaatteja ei tarvitse erotella blokkikohtaisiin ja muihin kovariaatteihin.

Olkoon $\mathbf{y} = (y_{11}, \dots, y_{nm})$ $n \times 1$ -kokoinen vaikutusten vektori ja \mathbf{X} $n \times (p+1)$ -kokoinen asetelmamatriisi. Suurimman uskottavuuden menetelmällä saadaan regressiokertoimille estimointiyhtälö (Konstantopoulos 2011, 67, 72)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}, \quad (11)$$

missä $V = \text{diag}(\sigma_{11}^2 + \tau^2, \dots, \sigma_{nm}^2 + \tau^2)$ on $n \times n$ -kokoinen vaikutusten kovarianssimatriisi. Kun samaistetaan $\sigma_i^2 \sim \sigma_{ij}^2$ kaikilla $i = 1, \dots, n$, vastaavat yhtälöt (5) ja (11) toisiaan; jälkimmäinen on vain muotoiltu eri tavalla matriisien avulla. Ainoa ero kaksitasoisen ja kolmitasoisen sekamallin estimoinnissa on heterogeenisuusparametrien estimointi, joka toki vaikuttaa lopulta välillisesti regressiokertoimien estimointiin. Heterogeenisuusparametrien τ^2 ja ω^2 estimoinnista on yksityiskohtaista tietoa Konstantopoulosin (2011) artikkelin sivuilla 72–74. Parametrien estimointi tapahtuu iteratiivisesti; eri parametreja päivitetään vuorotellen niin kauan, että riittävä tarkkuus saavutetaan.

3.2 Tilastollisesti riippuvat vaikutukset

Kappaleessa 3.1 vaikutusten on oletettu olevan toisistaan riippumattomia. Tietyissä mielessä tämä ei ole totta. Esimerkiksi mallin (2) oletuksista seuraa, että vaikutukset $y_i \sim N(\mu, \sigma_i^2 + \tau^2)$, missä odotusarvo μ ja toinen varianssikomponentti τ^2 ovat yhteisiä kaikille vaikutuksille. Tarkkaan ottaen vaikutukset ovat riippumattomia vain, kun ne on ehdollistettu kyseisillä estimoitavilla parametreilla. Tässä kappaleessa tämänkään ehdon ei oleteta toteutuvan; vaikutukset ovat *tilastollisesti riippuvia* toisistaan.

Vaikutukset voivat riippua toisistaan esimerkiksi yhteisen kontrolliryhmän tai useiden perättäisten mittauksen takia (Gleser & Olkin 2009, 358). Eri koeryhmien vertaaminen yhteiseen kontrolliryhmään säästää kustannuksia, mutta vaikutuksia ei voi enää käsitellä toisistaan riippumattomina. Samoin käy, jos vaikutusta halutaan mitata ajan funktiona useammalla kuin kahdella ajanhetkellä.

Gleser ja Olkin (2009) esittelevät metaregressiomallin estimointia, kun mallissa on vain kiinteitä vaikutuksia (ks. esim. Gleser & Olkin 2009, 3). Regressiokertoimien

estimaattorit saadaan kaavasta (11), kun $V = \text{diag}(V_1, \dots, V_m)$ on vaikutusten blokidiagonaalinen kovarianssimatriisi. Matriisit V_1, \dots, V_m eivät ole diagonaalimatriiseja, sillä vaikutukset eivät ole riippumattomia; diagonaalin ulkopuolella ovat vaikutusten kovarianssit $\text{Cov}(y_{ij}, y_{i'j}), i, i' = 1, \dots, n, j = 1, \dots, m$.

Kolmitasoinen sekamalli toisistaan riippuville vaikutuksille voidaan estimoida käyttämällä kiinteiden vaikutusten estimointiin jälleen yhtälöä (11), missä matriisin V diagonaalialkiot ovat $\sigma_{11}^2 + \tau^2, \dots, \sigma_{nm}^2 + \tau^2$ ja huomioimalla matriisin V kovarianssirakenne heterogeenisyysparametrien estimoinnissa. Konstantopoulos (2011, 73) olettaa matriisien V_1, \dots, V_m olevan diagonaalimatriiseja, mutta näin ei tarvitse olla. On syytä huomata, että V ei itsessään ole vaikutusten kovarianssimatriisi kolmitasoisessa sekamallissa (Konstantopoulos 2011, 72). Mallin estimointiin liittyvistä variansseista $\sigma_{11}^2, \dots, \sigma_{nm}^2$ ja kovariansseista kerrotaan kappaleessa 3.4.

3.3 Sekamallin ja kiinteiden vaikutusten mallin ero

Meta-analyysissä on periaatteessa mahdollista käyttää sekä kiinteiden vaikutusten mallia että sekamallia, mutta ei ole yhdentekevää, kumpi niistä valitaan (Hedges & Vevea, 1998). Jos tarkoitus on tehdä ainoastaan analyysissä mukana olevia tutkimuksia koskevaa ehdollista päättelyä, on kiinteiden vaikutusten malli sopiva. Kiinnostuksen kohteena on siis rajattu tutkimusten joukko, joka on havaittu kokonaan.

Otoksen ulkopuolelle jäänee kuitenkin aina joukko tutkimuksia, jotka on tehty, tullaan tekemään tulevaisuudessa tai jotka ovat jostakin syystä jääneet tekemättä. Hedges ja Vevea (1998) kehottavat käyttämään sekamallia, mikäli tarkoitus on ulottaa päätelmät koko tutkimusten perusjoukkoon mukaan lukien aineiston ulkopuolisiin tutkimuksiin. Meta-analyysissä, kuten tilastotieteessä yleensäkin, lienee tavanomaisempaa pyrkiä yleistämään tulokset koko populaatioon kuin tehdä ehdollista päättelyä käsillä olevalle aineistolle.

Yleinen harhaluulo on, että kiinteiden vaikutusten mallia voidaan käyttää täsmälleen silloin, kun todelliset vaikutukset ovat homogeenisiä. Näin ei kuitenkaan ole; keskeistä on, halutaanko tehdä ehdollista vai koko tutkimusten joukkoa koskevaa päätelyä. (Viechtbauer 2010, 4.)

3.4 Keskiarvojen vertailuun sopivia vaikutuksia

Tämän tutkielman aineistossa ryhmiä on vertailtu käyttäen keskiarvoja. Siksi seuraavaksi käydään läpi vaikutuksen estimaattoreja, jotka perustuvat keskiarvoihin. Nyt käsiteltävät vaikutukset ovat nimenomaan yksittäisiä vaikutuksen estimaatteja y_1, \dots, y_n , mutta yksinkertaisuuden vuoksi alaindeksit jätetään merkitsemättä. Käsiteltäessä sovellusaineiston meta-analyysin tuloksia kappaleessa 5 keskimääräistä vaikutusta merkitään kuin yksittäisiä vaikutuksia tässä kappaleessa.

3.4.1 Tilastollisesti riippumattomat vaikutukset

Oletetaan ensin vaikutusten olevan toisistaan riippumattomia. Borenstein ym. (2009, 21–32) esittävät kolme keskiarvoihin pohjautuvaa vaikutusta, joista yksinkertaisin on *keskiarvojen erotus* (Borenstein ym. 2009, 21–22). Olkoot μ_T koeryhmän T ja μ_C kontrolliryhmän C keskiarvo sekä σ_T ja σ_C niitä vastaavat keskihajonnat. Olkoot vielä $\hat{\mu}_T$ koeryhmän ja $\hat{\mu}_C$ kontrolliryhmän otoskeskiarvo. Tällöin koko populaation *keskiarvojen erotusta*

$$D = \mu_T - \mu_C$$

voidaan estimoida otoskeskiarvojen erotuksella

$$\hat{D} = \hat{\mu}_T - \hat{\mu}_C.$$

Olkoot lisäksi s_T ja s_C otoskeskiarvojen keskihajonnat sekä n_T ja n_C ryhmäkohtaiset otoskoot. Silloin erotukselle \hat{D} voidaan helposti johtaa varianssiestimaattori

$$\hat{\sigma}_D^2 = \frac{s_T^2}{n_T} + \frac{s_C^2}{n_C}. \quad (12)$$

Erotus D käy vaikutukseksi, jos tutkittavaa ominaisuutta on mitattu käyttäen aina samaa mittayksikköä. Mikäli mittayksikkö on hyvin tunnettu, meta-analyysin tulokset on helppo tulkita. Jos esimerkiksi koeryhmän keskimääräinen fyysinen aktiivisuus on 9500 askelta päivässä ja kontrolliryhmän 7500 askelta päivässä, ovat koeryhmän osallistujat olleet keskimäärin 2000 askelta päivässä enemmän liikkeellä. Tutkielman aineistossa on kuitenkin useita eri mittayksiköitä, jotka eivät ole keskenään vertailukelpoisia. Siksi keskiarvojen erotusta ei pystytä käyttämään. Sen sijaan voidaan hyödyntää esimerkiksi *standardoitua keskiarvojen erotusta* (Borenstein ym. 2009, 25–27), johon tarvitaan yhdistetty keskihajonta

$$s = \sqrt{\frac{(n_T - 1) s_T^2 + (n_C - 1) s_C^2}{n_T + n_C - 2}}. \quad (13)$$

Yhdistetyn keskihajonnan ja otoskeskiarvojen avulla voidaan laskea standardoitu otoskeskiarvojen erotus

$$\hat{d} = \frac{\hat{\mu}_T - \hat{\mu}_C}{s}, \quad (14)$$

jolla estimoidaan koko populaation standardoitua keskiarvojen erotusta

$$d = \frac{\mu_T - \mu_C}{\sigma},$$

missä $\sigma = \sigma_1 = \sigma_2$. Toisin kuin keskiarvojen erotuksen yhteydessä, nyt siis populaatiokeskihajonnat σ_1 ja σ_2 oletetaan yhtä suuriksi. Erotuksen \hat{d} varianssille saadaan arvio

$$\hat{\sigma}_d^2 = \frac{n_T + n_C}{n_T n_C} + \frac{\hat{d}^2}{2(n_T + n_C)}. \quad (15)$$

Koska keskihajonnan yksikkö on sama kuin alkuperäisen muuttujan yksikkö, standardoidulla keskiarvojen erotuksella ei ole yksikköä lainkaan. Tämä antaa tilaisuuden yhdistää eri mittayksiköitä toisiinsa. Toisaalta keskiarvojen erotuksen tarjoama

tulkinta menetetään. Siksi tarkastellaan vielä *keskiarvojen osamäärää* (Borenstein ym. 2009, 30–31), joka on toinen vaihtoehto erilaisten tulosmuuttujien yhdistämiseen. Populaatiotasolla keskiarvojen osamäärä on

$$R = \frac{\mu_T}{\mu_C}$$

ja sitä estimoidaan otoskeskiarvojen osamäärällä

$$\hat{R} = \frac{\hat{\mu}_T}{\hat{\mu}_C}. \quad (16)$$

Varianssin estimoimiseksi tehdään logaritmimuunnos

$$\begin{aligned} \log(\hat{R}) &= \log\left(\frac{\hat{\mu}_T}{\hat{\mu}_C}\right) \\ &= \log(\hat{\mu}_T) - \log(\hat{\mu}_C). \end{aligned}$$

Hedges, Gurevitch ja Curtis (1999) esittävät varianssille approksimaation

$$\hat{\sigma}_{\log(\hat{R})}^2 = \frac{s_T^2}{n_T \hat{\mu}_T^2} + \frac{s_C^2}{n_C \hat{\mu}_C^2}, \quad (17)$$

joka voidaan johtaa käyttämällä deltamenetelmää (ks. Ver Hoef 2012, 124). Merkitään $f(x) = \log(x)$, jolloin $f'(x) = \frac{1}{x}$. Ryhmien välisestä riippumattomuudesta ja deltamenetelmästä seuraa, että populaatiotason osamäärän varianssi

$$\begin{aligned} \sigma_{\log(R)}^2 &= \text{Var}(\log(\mu_T) - \log(\mu_C)) \\ &= \text{Var}(\log(\mu_T)) + \text{Var}(\log(\mu_C)) \\ &\approx f'(\mu_T)^2 \text{Var}(\mu_T) + f'(\mu_C)^2 \text{Var}(\mu_C) \\ &= \frac{1}{\mu_T^2} \frac{\sigma_T^2}{n_T} + \frac{1}{\mu_C^2} \frac{\sigma_C^2}{n_C}. \end{aligned}$$

Kaavan (17) mukainen approksimaatio saadaan sijoittamalla populaatioparametrien paikalle vastaavat otoksista lasketut tunnusluvut.

Keskiarvojen osamäärää käytettäessä mallin estimointi tehdään logaritmiasteikolla, mutta lopuksi tulokset muunnetaan osamääräksi eksponenttimuunnoksella. Toisin kuin standardoidulla keskiarvojen erotuksella, keskiarvojen osamäärällä on havainnollinen tulkinta. Keskiarvojen erotuksen kohdalla esitetystä esimerkkitapauksessa osamäärä on $\frac{9500 \text{ askelta/päivä}}{7500 \text{ askelta/päivä}} \approx 1.27$. Koeryhmän osallistujat ovat siis olleet 1.27 kertaa aktiivisempia kuin kontrolliryhmän osallistujat. Yhtäpitävästi voidaan sanoa, että koeryhmän osallistujat ovat liikkuneet keskimäärin 27 % enemmän. Ykköistä pienempi osamäärä on mahdollista tulkita esimerkiksi edelliseen tapaan laskemalla sen käänteisluku ja vaihtamalla ryhmien roolit päikseen. Keskiarvojen osamäärää käytettäessä muuttujien on oltava suhdeasteikollisia. Sovellusaineistossa fyysistä on mitattu aina suhdeasteikollisella muuttujalla, joten keskiarvojen osamäärä kelpaa sen analysointiin.

3.4.2 Yhteisestä kontrolliryhmästä aiheutuva riippuvuus

Jos kahden tai useamman vaikutuksen estimoinnissa käytetään yhteistä kontrolliryhmää, vaikutukset riippuvat tilastollisesti toisistaan. Olkoot T_1, \dots, T_k koeryhmiä otoskokoineen n_{T_1}, \dots, n_{T_k} siten, että koeryhmillä on yhteinen kontrolliryhmä C. Tällöin kaikilla $i = 1, \dots, k$ koeryhmää T_i vastaavan standardoidun keskiarvojen erotuksen varianssilla on estimaattori

$$\hat{\sigma}_{\hat{d}_{T_i}}^2 = \frac{n_T + n_C}{n_T n_C} + \frac{\hat{d}^2}{2n_{tot}}, \quad (18)$$

missä $n_{tot} = n_C + \sum_{i=1}^k n_{T_i}$ (Gleser & Olkin 2009, 365). Koeryhmiä T_i ja T_j vastaavien vaikutusten kovarianssin estimaattori on (Gleser & Olkin 2009, 365)

$$\widehat{\text{Cov}}(\hat{d}_{T_i}, \hat{d}_{T_j}) = \frac{n_T + n_C}{n_T n_C} + \frac{\hat{d}_{T_i} \hat{d}_{T_j}}{2n_{tot}}. \quad (19)$$

Keskiarvojen osamäärän logaritmillemme $\log(\hat{R})$ varianssiestimaattori on

$$\hat{\sigma}_{\log(\hat{R}_{T_i})}^2 = \frac{s_{T_i}^2}{n_{T_i} \hat{\mu}_{T_i}^2} + \frac{s_C^2}{n_C \hat{\mu}_C^2} \quad (20)$$

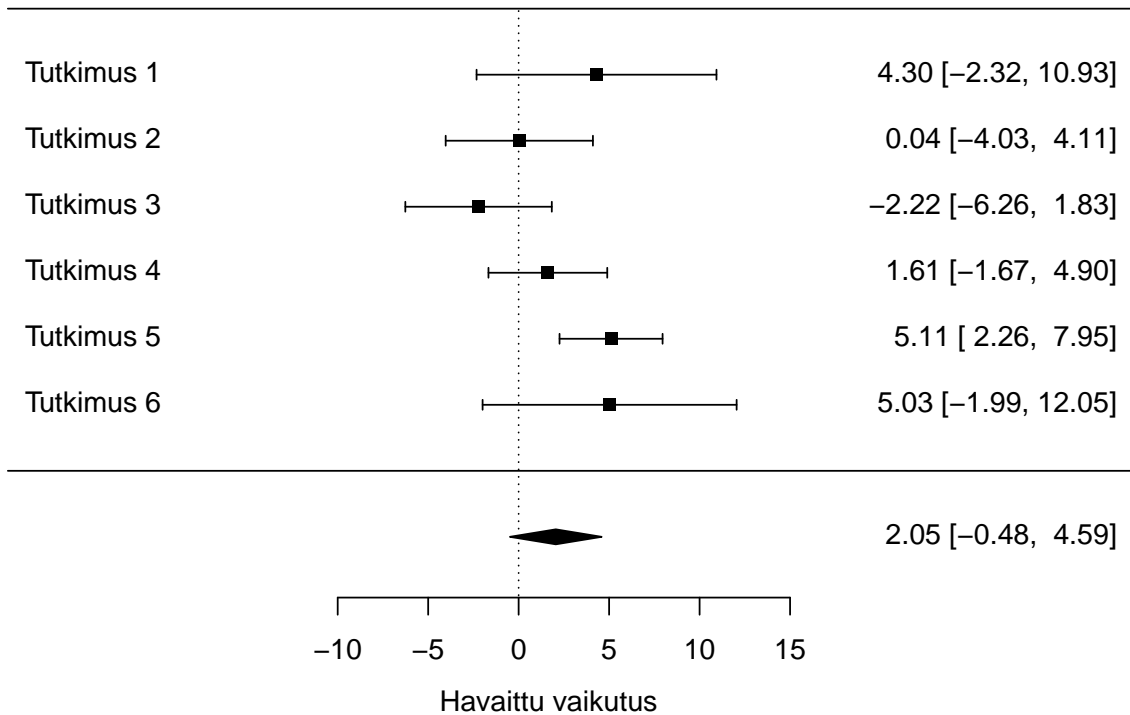
ja kovarianssiestimaattori

$$\widehat{\text{Cov}}(\log(\hat{R}_{T_i}), \log(\hat{R}_{T_j})) = \frac{s_C^2}{n_C \hat{\mu}_C^2} \quad (21)$$

(Lajeunesse 2011, 2051).

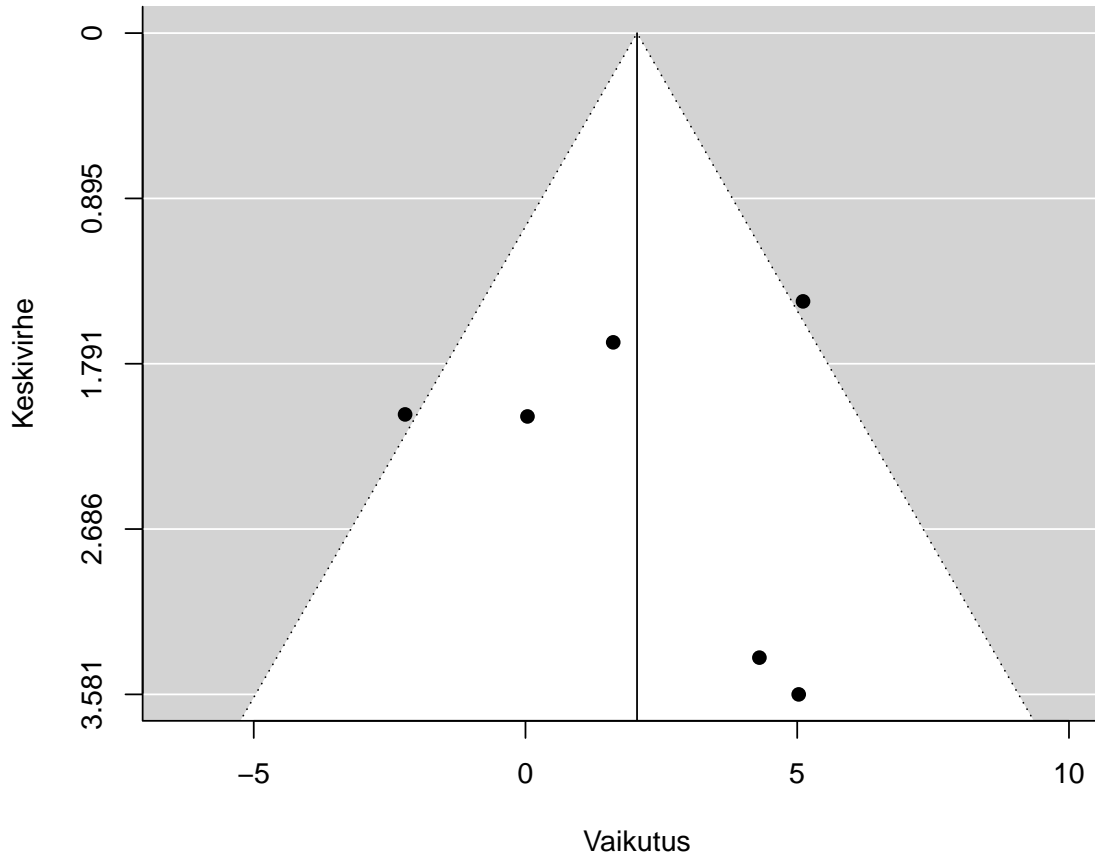
3.5 Graafiset tarkastelut ja julkaisuharha

Meta-analyysin tuloksia esitetään usein myös graafisesti *metsäkuvioiden* ja *suppilo-kuvioiden* avulla (ks. esimerkkikuvat 3 ja 4). Metsäkuvioista on eniten apua tulkittaessa mallia, jossa ei ole kovariaatteja. Yksinkertaisimmillaan metsäkuvio sisältää kunkin tutkimuksen nimen, vaikutuksen estimaatin luottamusväleineen ja niin sanotun salmiakkikuvion, josta nähdään keskimääräisen vaikutuksen estimaatti ja sen luottamusväli.



Kuva 3: Simuloidun aineiston metsäkuvio. Keskimääräisen vaikutuksen estimaatti on positiivinen mutta ei poikkea tilastollisesti merkitsevästi nolasta.

Suppilokuvio on hajontakuvio, jossa vaaka-akselilla on vaikutuksen estimaatti ja pystyakselilla jokin vaikutuksen tarkkuuden mitta, esimerkiksi keskivirhe. Tarkalleen ottaen keskivirhe on epätarkkuuden mitta, joten keskivirhettä käytettäessä pystyakseli käännetään ylösalaisin. Hajontakuvioon lisätään vielä janat, jotka yhdessä vaaka-akselin kanssa muistuttavat suppiloa. Janat määräävät vaakasuunnassa 95 %:n luottamusvälin vaikutuksen estimaatille, kun keskivirhe on kiinnitetty. Jos moni piste sijaitsee suppilon ulkopuolella, vaikutukset ovat heterogeenisiä.



Kuva 4: Simuloidun aineiston suppilokuvio. Silmämääräisesti vaikutukset ovat hie- man heterogeenisiä, eikä julkaisuharhaakaan voida sulkea pois.

Suppilokuvioilla on toinenkin käyttötarkoitus. Mahdollinen valikoiva tutkimuksien julkaisu saattaa aiheuttaa harhaa, niin kutsuttua julkaisuharhaa, meta-analyysin tuloksiin. Julkaisuharhan lähteitä on monenlaisia (Borenstein ym. 2009, 277–281). Harha voi liittyä systemaattisen kirjallisuuskatsauksen toteutukseen; englanninkie- liset tutkimukset saatetaan löytää muita todennäköisemmin (kieliharha), samoin ilmaiseksi saatavilla olevat (kustannusharha) ja paljon viitatus tutkimukset (viit- tausharha).

Usein julkaisuharha kytkeytyy julkaisukäytäntöihin. Tutkimusten raportointi voi ol- la valikoivaa; tilastollisesti merkitsevät tulokset julkaistaneen melko varmasti, mutta muuten tutkimus voi jäädä niin sanotusti pöytälaatikkoon. Jos ei-merkitseviin tu- loksiin liittyy vielä pieni otoskoko, on julkaisematta jättämisen riski suuri. Suuret tutkimukset sen sijaan yleensä julkaistaan tuloksista huolimatta, sillä niiden eteen on nähty paljon vaivaa. (Borenstein ym. 2009, 278, 281.) Tämän tyyppinen harha saa aikaan tyhjän tilan suppilokuvion vasempaan alakulmaan (Borenstein ym. 2009, 283).

Vaikutusten ja tutkimusten koon yhteyttä voidaan tarkastella myös tilastollisil- la testeillä. Begg ja Mazumdar (1994) ovat luoneet Kendallin τ^1 :hun perustuvan

¹Korrelaatiokerrointa τ ei pidä sekoittaa heterogeenisyyteen τ^2 .

järjestyskorrelaatiotestin, joka mittaa tietyllä tavalla standardoitujen vaikutusten ja vaikutusten varianssien korrelaatiota. Jos korrelaatiokerroin poikkeaa merkitsevästi nolosta, aineistossa saattaa olla julkaisuharhaa; tavallisin tapaus lienee positiivinen korrelaatio, joka voi olla seurausta yllä mainitusta valikoitumisesta. Järjestyskorrelaatiotesti on joissain tilanteissa tehoton (Begg & Mazumdar 1994, 1098), joten ei-merkitsevää p-arvoa ei aina kannata tulkita varmaksi merkiksi harhattomuudesta.

Jos pieniä tutkimuksia näyttää puuttuvan, voidaan julkaisuharhan vaikutusta hallita rajoittamalla meta-analyysi vain riittävän suuriin tutkimuksiin. Ei ole kuitenkaan olemassa yksikäsitteistä määritelmää, milloin tutkimukset ovat riittävän suuria. (Borenstein ym. 2009, 287.) Lisäksi on huomattava, että epäsymmetrinen suppilokuvio ei ole aina seurausta julkaisuharhasta. Joskus pienillä tutkimuksilla on ominaispiirteitä, joiden ansiosta niiden vaikutukset ovat oikeasti suuria (Borenstein ym. 2009, 291, ks. myös Sutton 2009, 438–440).

4 Sovellusaineiston meta-analyysin toteutus

Sovellusaineiston meta-analyysi tehdään erikseen loppuvaiheelle ja seurantavaiheelle hyödyntäen kolmitasoista sekamallia. Joidenkin vaikutusten laskemisessa on käytetty yhteistä kontrolliryhmää, mistä aiheutuva vaikutusten välinen riippuvuus sisällytetään meta-analyysiin kappaleen 3.2 mukaisesti. Riippuvuuden merkityksen arvioimiseksi mallit sovitetaan myös olettaen vaikutukset riippumattomiksi (ks. kappale 3.1). Kolmitasoisen sekamallin blokit muodostavat pääsääntöisesti vaikutukset, joilla on yhteinen kontrolliryhmä. Ainoat poikkeukset ovat Nakaden ym. (2012) kaksi vaikutusta ja Demark-Vahnefriedin ym. (2014) neljä vaikutusta, joita molempia kohdellaan yhtenä blokkina. Vaikutuksena käytetään kaavan (16) keskiarvojen osamäärää. Lisäksi kokeillaan kaavan (14) standardoitua keskiarvojen erotusta, jotta sen toimivuutta voidaan verrata keskiarvojen osamäärään.

Yhteensä 28 kovariaatista 26:ta (kaikkia paitsi *durfollow* ja *lengthfollow*) käytetään loppuvaiheen metaregressiossa ja 25:tä (kaikkia paitsi *intensity*, *durint* ja *lengthint*) seurantavaiheen metaregressiossa. Tutkimusten vaihtelevan raportoinnin vuoksi osa kovariaattien arvoista puuttuu. Kaikkia motivoituneita koskevien kovariaattien arvot tunnetaan, mutta varsinkin seurantavaiheen kovariaatit (*durfollow* ja *lengthfollow*) sekä koeryhmän kontaktiaika minuutteina *intervention* aikana (*intensity*) tiedetään suhteellisen harvoin. Puuttuvuuden oletetaan olevan satunnaista, ja mallit sovitetaan aineiston osaan, jossa mallissa mukana olevat kovariaatit on havaittu.

Suurimmassa osassa tutkimuksia ryhmäkeskiarvot ja niiden keskihajonnat sekä otokoot on raportoitu asianmukaisesti. Joissain tapauksissa on ilmoitettu keskihajontojen sijaan keskivirheet tai luottamusvälit, joita käyttäen keskihajonnat on voitu laskea. Jos haluttujen tunnuslukujen tai niiden kanssa yhtäpitävien tietojen sijaan tarjolla on muita keski- ja hajontalukuja, niitä hyödynnetään olettamalla normaali-jakautuneisuus; esimerkiksi keskiarvo korvataan mediaanilla ja kvartiiliväli samais-

tetaan havaintojen (ei niiden keskiarvon) 50 %:n luottamusvälin kanssa.

Roesch ym. (2010) ja Broekhuizen ym. (2012) käsitellään muista tutkimuksista poikkeavasti. Roeschin (2012) ym. tulokset ovat valmiiksi logaritmiasteikolla², joten keskiarvojen osamäärää käytettäessä vaikutus on yksinkertaisesti ryhmäkeskiarvojen erotus ja sen varianssi lasketaan kuten keskiarvojen erotukselle kaavan (12) mukaan. Broekhuizenin ym. (2012) kohdalla toimitaan muuten samoin, mutta raportoinnissa tapahtunut virhe³ otetaan huomioon tekemällä ryhmäkeskiarvoille logaritmimuunnos. Tilanne kääntyy pääläelleen, kun halutaan käyttää standardoitua keskiarvojen erotusta; Roeschin ym. (2012) ryhmäkeskiarvoille tehdään eksponenttimuunnos ja muunnetuille ryhmäkeskiarvoille lasketaan varianssit deltamenetelmällä⁴. Toisin kuin kaavan (17) johtamisessa muunnosfunktio on logaritmfunktion sijaan eksponenttifunktio $f(x) = e^x = f'(x)$.

Metaregressiomallin valinta tehdään kahdessa vaiheessa. Ensin sovitetaan mallit, joissa on mukana yksi kovariaatti kerrallaan. Jos kovariaatin merkitsevyyttä mittaava p-arvo on suurempi kuin 0.05, kovariaatti hylätään. Mikäli kaikki kovariaatit hylätään, valitaan niin sanottu nollamalli ilman kovariaatteja. Muutoin mallinvalinnan toiseen vaiheeseen pääsevät kovariaatit, joita ei edellä hylätty. Hylkäämättömät kovariaatit järjestetään ensimmäisen vaiheen p-arvojen mukaan pienimmästä alkaen. Toisessa vaiheessa lisätään nollamalliin kovariaatteja tässä järjestyksessä kunnes malli ei uskottavuusosamäärän testin perusteella parane ($p \geq 0.05$) tai malliin on lisätty kaikki kovariaatit, joita ei ensimmäisessä vaiheessa hylätty. Malleista viimeinen, joka on edellistä merkitsevästi parempi, valitaan.

Mallit sovitetaan R-ympäristössä (R Core Team 2017) meta-analyysiin tarkoitettun *metafor*-paketin avulla (Viechtbauer 2010). Suurin osa meta-analyysistä tehdään suoraan *metafor*-paketin avulla. Vaikutusten estimaatit sekä niiden varianssit ja kovarianssit on laskettu ilman paketin ominaisuuksia. Liitteessä 1 on esimerkki analyysissä käytetystä koodista.

5 Meta-analyysin tulokset sovellusaineistossa

5.1 Loppuvaihe

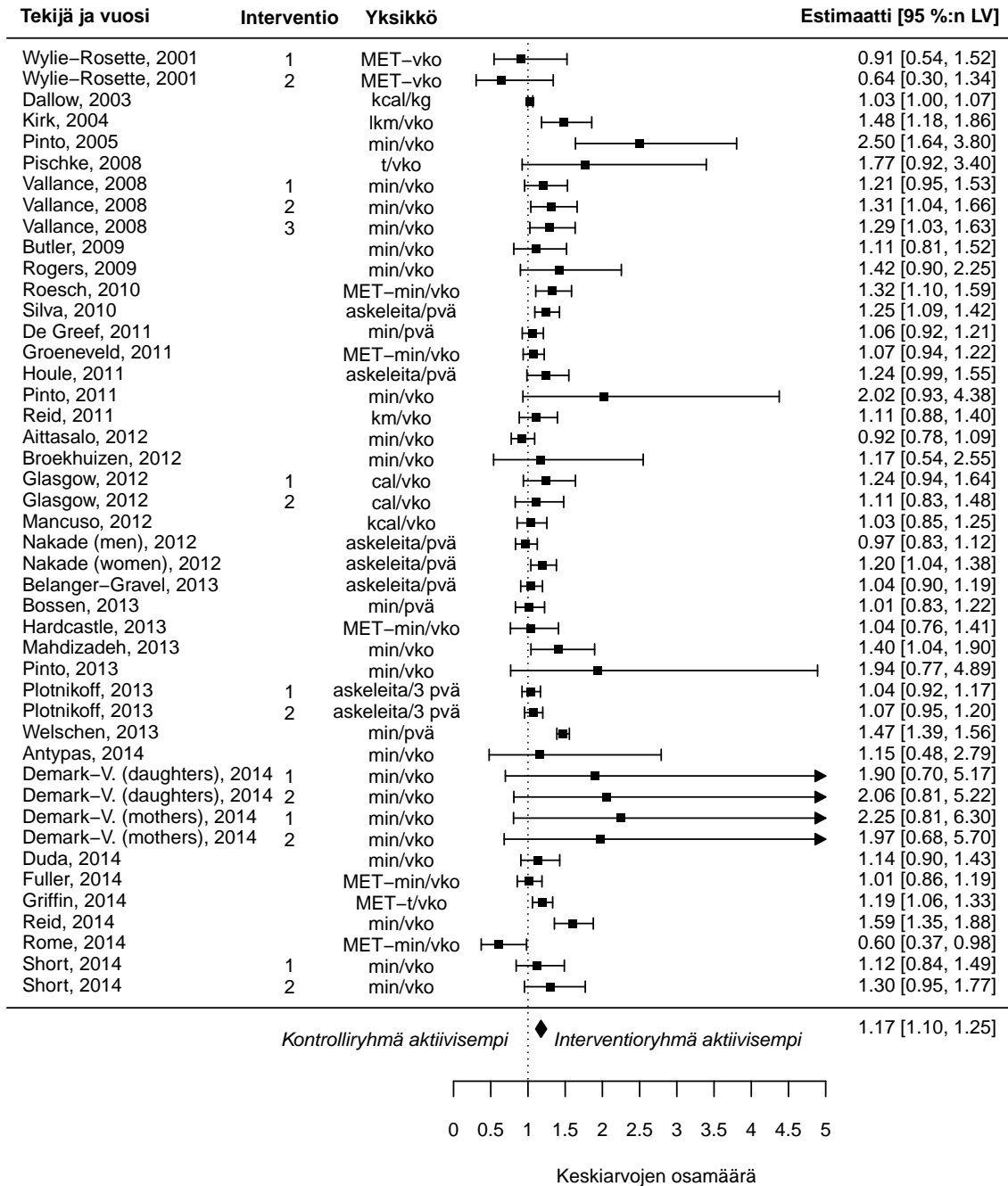
Koe- ja kontrolliryhmä poikkeavat toisistaan loppuvaiheessa; fyysisessä aktiivisuudessa on merkitsevä ero koeryhmän hyväksi ($\hat{R} = 1.174$, $p = 3 \cdot 10^{-7}$, 95 %:n luot-

²Roesch ym. (2010) tekivät fyysisen aktiivisuuden arvoille ensin logaritmimuunnoksen ja laskivat sitten aritmeettiset ryhmäkeskiarvot; tämä johtaa saamaan lopputulokseen kuin jos lasketaan ensin geometriset ryhmäkeskiarvot, joille tehdään logaritmimuunnos. Aritmeettinen keskiarvo ja geometrinen keskiarvo eivät tietenkään ole samat, mutta tähän pulmaan ei perehdytä tässä työssä.

³Broekhuizen ym. (2012) raportoivat vahingossa geometriset ryhmäkeskiarvot ja geometrinen ryhmäkeskiarvojen *logaritmien* keskihajonnat.

⁴Broekhuizenin ym. (2012) tapauksessa varianssit lasketaan samalla tavalla, mutta ryhmäkeskiarvoille ei tehdä eksponenttimuunnosta.

tamuväli [1.104, 1.247], ks. myös kuva 5). Koeryhmän osallistujat ovat liikkuneet arviolta keskimäärin 17 % enemmän kuin kontrolliryhmän osallistujat. Vaikutukset ovat heterogeenisiä ($p = 2 \cdot 10^{-21}$), joten vaikutusten keskinäiset erot eivät selity pelkällä satunnaisvaihtelulla.



Kuva 5: Metsäkuvio loppuvaiheen meta-analyysistä ilman kovariaatteja. Saman tutkimuksen eri koeryhmien interventioihin perustuvat vertailut on eroteltu toisistaan numeroin.

Alustavasti loppuvaiheen metaregression ensimmäisessä vaiheessa ainoastaan tut-

kimuksen laatu (*quality*) on merkitsevä ($p = 0.035$). Toisaalta kuitenkin huomaetaan, että kontrolliryhmän tyyppin (*control*) ja keskimääräisen vaikutuksen yhteys näyttää likimain lineaariselta. Jatkovana kontrolliryhmän tyyppi onkin merkitsevä ($p = 0.036$). Kun oletetaan malli, jossa on edelliset kaksi kovariaattia, saadaan

$$\begin{aligned}\hat{R} &= \exp(0.1231 + 0.0348 \cdot \textit{quality} - 0.0667 \cdot \textit{control}) \\ &= \exp(0.1231) \cdot \exp(0.0348 \cdot \textit{quality}) \cdot \exp(-0.0667 \cdot \textit{control}) \\ &= \exp(0.1231) \cdot \exp(0.0348)^{\textit{quality}} \cdot \exp(-0.0667)^{\textit{control}} \\ &\approx 1.131 \cdot 1.035^{\textit{quality}} \cdot 0.935^{\textit{control}}.\end{aligned}$$

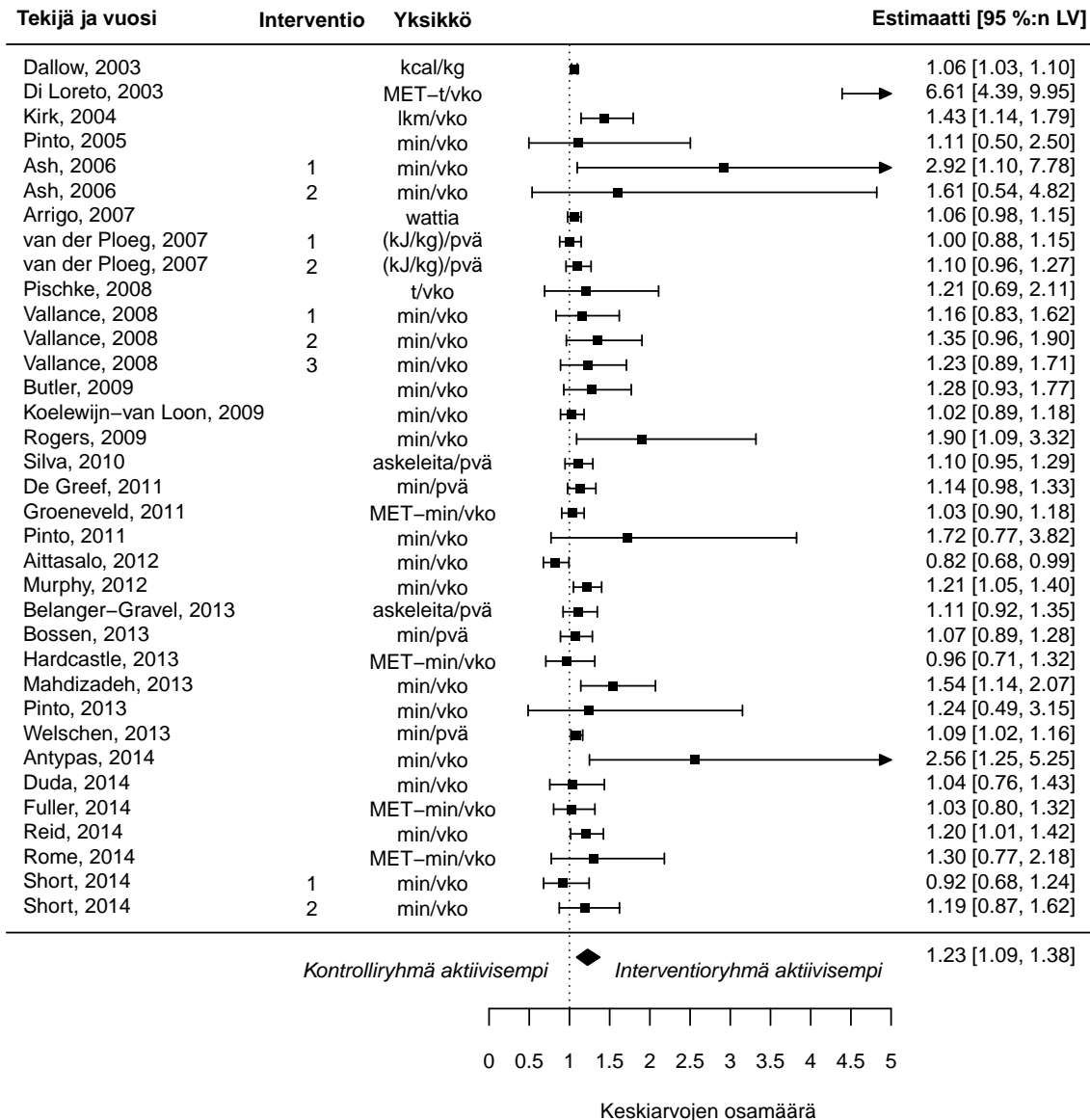
Tämä kahden kovariaatin malli valitaan, sillä uskottavuusosamäärän testin mukaan se on merkitsevästi parempi kuin malli, jossa on yksistään laatu ($p = 0.018$). Tutkimuksen laadun kertoimen luottamusväli on $[1.007, 1.064]$ ja p-arvo 0.013. Kontrolliryhmän tyyppiä vastaavan kertoimen luottamusväli on $[0.886, 0.987]$ ja p-arvo 0.015.

Laadun kerroin on suurempi kuin yksi, joten vaikutus riippuu positiivisesti tutkimuksen laadusta. Jos kahdessa tutkimuksessa kontrolliryhmä on saanut samankaltaisen ohjeistuksen mutta tutkimusten laadussa on yhden pisteen ero, laadukkaamman tutkimuksen ennustettu vaikutus on 3.5 % suurempi. Riippuvuuden luonne kontrolliryhmän ohjeistuksen ja vaikutuksen välillä on päinvastainen. Mikäli kaksi tutkimusta ovat laadultaan identtisiä mutta toisessa kontrolliryhmän ohjeistus on luokkaa suurempi, on tämän ennustettu vaikutus 6.5 % pienempi.

Loppu- ja seurantavaiheen yhdistetyssä aineistossa laatupisteet vaihtelevat välillä 2–8. Esimerkiksi jos koehenkilölle on annettu ohjeistus fyysiseen aktiivisuuteen ja tutkimuksen laatupisteet vaihtelevat 2–8, vaihtelee osamäärän estimaatti välillä $[1.131 \cdot 1.035^2 \cdot 0.935^3, 1.31 \cdot 1.035^8 \cdot 0.935^3] \approx [0.990, 1.217]$. Mallin ennusteisiin kannattaa suhtautua varauksella; todellisuudessa on hyvin epätodennäköistä, että koeryhmän interventio olisi kontrolliryhmän interventiota huonompi. Vaikka malli selittää 43 % heterogeenisyydestä, jäännösheterogeenisuus on merkitsevää ($p = 2 \cdot 10^{-6}$).

5.2 Seurantavaihe

Tulokset ilman kovariaatteja ovat seurantavaiheessa melko samankaltaiset kuin loppuvaiheessa. Osamäärän estimaatti on $\hat{R} = 1.225$ ($p = 7 \cdot 10^{-4}$, 95 %:n luottamusväli $[1.089, 1.378]$, ks. kuva 6). Ryhmien välinen ero fyysisessä aktiivisuudessa on noin 23 %, jopa suurempi kuin loppuvaiheessa. Täälläkin kertaa vaikutukset ovat heterogeenisia ($p = 2 \cdot 10^{-12}$). Seurantavaiheessa kaikki kovariaatit hylätään ensimmäisessä vaiheessa, joten meta-analyysimalliin ei tule yhtään kovariaattia. Siten heterogeensyyden lähde jää tunnistamatta.



Kuva 6: Metsäkuvio seurantavaiheen meta-analyysistä ilman kovariaatteja.

5.3 Poikkeavat havainnot

Useimmat vaikutusten estimaatit luottamusväleinen ovat sekä loppu- että seurantavaiheessa järkevässä mittakaavassa. Kuvasta 6 kuitenkin nähdään Di Loreton ym. (2003) poikkeavan huomattavasti muista seurantavaiheesta. Tutkimuksen tekijät ovat vahvistaneet luvut oikeiksi, mutta siitä huolimatta analyysi on tehty myös ilman kyseistä tutkimusta. Muutos on selkeä; osamäärän estimaatiksi saadaan $\hat{R} = 1.078$ ($p = 2 \cdot 10^{-10}$, 95 %:n luottamusväli [1.054, 1.104]). Fyysisen aktiivisuuden ero ryhmien kesken olisi vain 7.8 %, mikä on pienempi kuin loppuvaiheessa ja selkeästi pienempi kuin seurantavaiheessa Di Loreto ym. (2003) mukaan otettuna. Vaikutukset ovat yhä heterogeenisiä, mutta p-arvo on selvästi suurempi ($p = 0.026$).

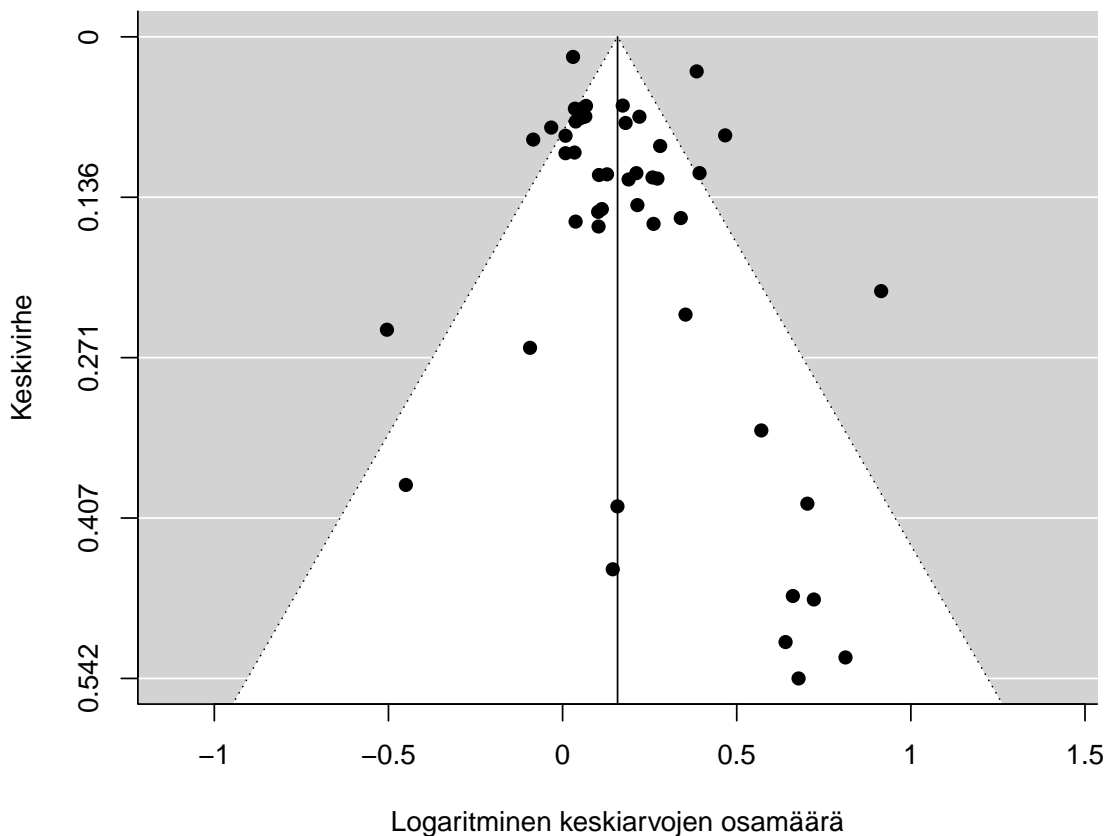
Di Loreton ym. (2003) pois jättämisen seurauksena myös seurantavaiheen meta-regression tulokset muuttuvat. Aiemmin merkitseviä kovariaatteja ei ollut, mutta nyt malliksi valikoituu

$$\begin{aligned}\hat{R} &= \exp(0.0620 + 0.0706 \cdot multispec) \\ &\approx 1.064 \cdot 1.073^{multispec}.\end{aligned}$$

Vakiotermin 1.064 ($p < 0.001$, 95 %:n luottamusväli [1.037, 1.092]) ja regressiokerroin 1.073 ($p = 0.016$, 95 %:n luottamusväli [1.013, 1.137]) ovat merkitseviä. Mallin mukaan koe- ja kontrolliryhmän ero fyysisessä aktiivisuudessa on 6.4 %, jos koeryhmän intervention on antanut yksi asiantuntija. Ero on 14 %, kun asiantuntijoita on useita. Kovariaatti selittää jopa 75 % heterogeenisyydestä, eikä jäännösheterogeenisuus ole enää aivan tilastollisesti merkitsevää ($p = 0.068$).

5.4 Julkaisuharhan arviointi ja huomioiminen

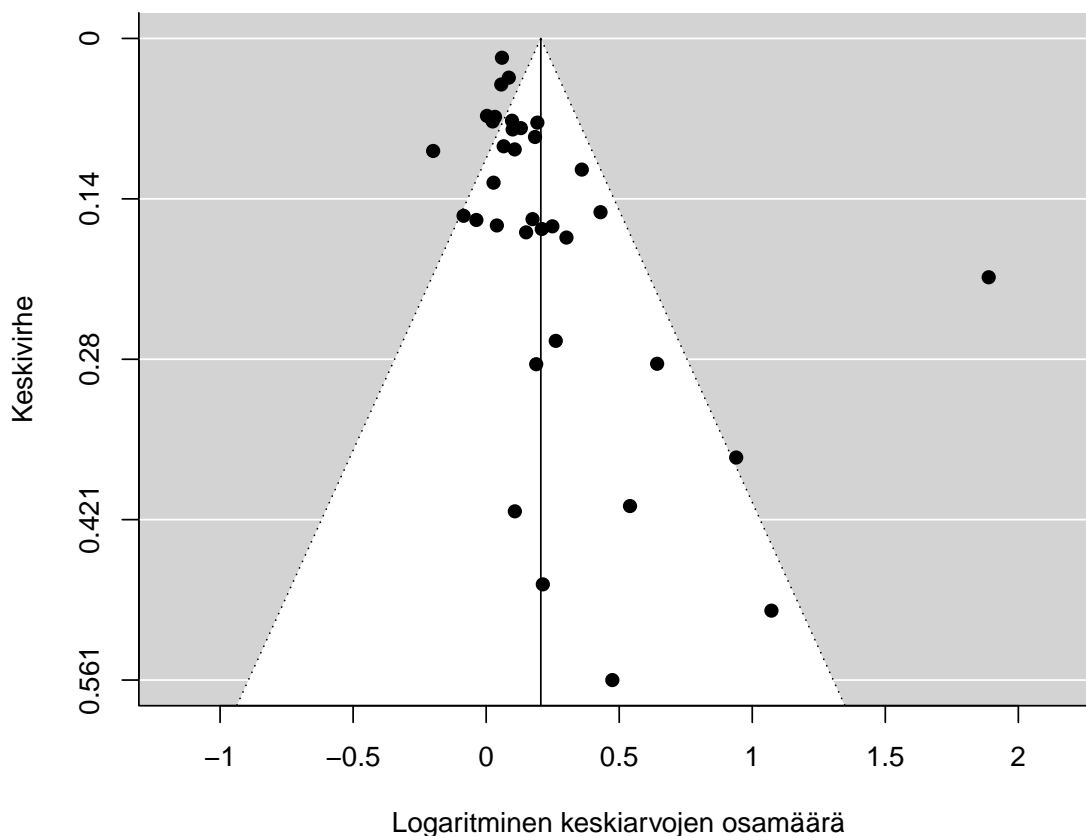
Aineistossa on viitteitä julkaisuharhasta. Loppuvaiheen suppilokuviossa (kuva 7) on paljon tyhjää tilaa vasemmassa alakulmassa. Pieniä tutkimuksia, joiden logaritminen keskiarvojen osamäärä on negatiivinen tai hyvin lähellä nollaa, on saattanut jäädä meta-analyysin ulkopuolelle. Järjestyskorrelaatiotestin tulos ei ole merkitsevää (Kendallin $\tau = 0.186$, $p = 0.073$). Testin tehottomuus huomioiden julkaisuharhaa ei kuitenkaan voi sulkea pois.



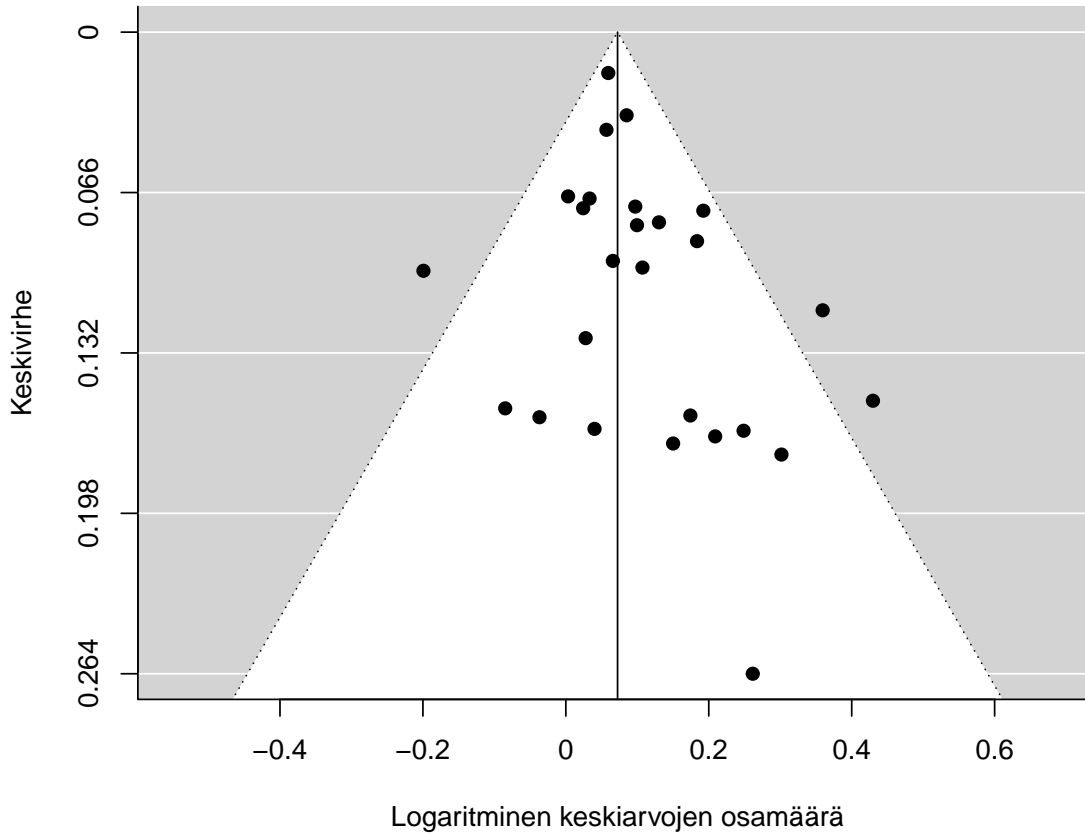
Kuva 7: Suppilokuva lopputulosten analyysistä ilman kovariaatteja.

Seurantavaiheessa epäily julkaisuharhasta on vieläkin vahvempi. Tulkinta ei juuri muutu, onpa Di Loreto ym. (2003) mukana (kuva 8) tai ei (kuva 9), joskin ilman Di Loretoa ym. (2003) vaikutusten jakauma keskimääräisen vaikutuksen ympärillä näyttää hiukan enemmän normaalijakautuneelta. Järjestyskorrelaatiotestin tulos on merkitsevä sekä Di Loreton ym. (2003) kanssa ($\tau = 0.345$, $p = 0.003$) että ilman sitä ($\tau = 0.323$, $p = 0.007$).

Koska julkaisuharhan olemassaoloa ei voida sulkea pois, on se otettava huomioon herkkyysanalyysillä. Kuvaa 9 katsomalla näyttää järkevältä poistaa seurantavaiheen analyysistä poikkeavan Di Loreton ym. (2003) lisäksi yhdeksän muuta keskivirheeltään suurta vaikutusta; tällöin suppilokuvio olisi likimain symmetrinen. Vastaavalla päättelyllä päädytään siihen, että kuusi keskivirheeltään suurta vaikutusta olisi syytä jättää pois. Jos samasta tutkimuksesta on monta vaikutusta ja yksi vähintään yksi niistä on epäilyksenalainen, on luonnollista poistaa kaikki kyseisen tutkimuksen vaikutukset. Jos taas vaikutus poistetaan seurantavaiheen (tai loppuvaiheen) analyysistä, on luontevaa poistaa se myös loppuvaiheesta (tai seurantavaiheesta).



Kuva 8: Suppilokuvio seurantavaiheen analyysistä ilman kovariaatteja.



Kuva 9: Suppilokuvio seurantavaiheen analyysistä ilman kovariaatteja, kun Di Loreto ym. (2003) on poistettu.

Yllä olevan järjestyksellä kymmenen loppuvaiheen vaikutusta ja yhdeksän seurantavaiheen vaikutusta jätetään pois. Molemmista vaiheista poistettavat tutkimukset ovat Antypas ja Wangberg (2014), Pinto, Frierson, Rabin, Trunzo ja Marcus (2005), Pinto, Papandonatos ja Goldstein (2011), Pinto, Papandonatos, Goldstein, Marcus ja Farrell (2013), Pischke, Scherwitz, Weidner ja Ornish (2008), Rogers ym. (2009). Lisäksi loppuvaiheesta jätetään pois Demark-Vahnefried ym. (2014) (neljä vaikutusta) sekä seurantavaiheesta Ash ym. (2006) (kaksi vaikutusta) ja Di Loreto ym. (2003).

Odotetusti keskimääräisen vaikutuksen estimaatit kutistuvat kohti ykköstä, mutta eivät kovin paljoa. Loppuvaiheessa $\hat{R} = 1.143$ ($p = 7 \cdot 10^{-6}$, luottamusväli [1.078, 1.211]) ja seurantavaiheessa $\hat{R} = 1.075$ ($p = 10^{-9}$, luottamusväli [1.050, 1.100]). Ryhmien välinen suhteellinen ero pienenee loppuvaiheessa 17 prosentista 14 prosenttiin ja seurantavaiheessa 7.8 prosentista 7.5 prosenttiin. Loppuvaiheessa jäljelle jääneet vaikutukset ovat edelleen selkeästi heterogeenisiä ($p = 8 \cdot 10^{-21}$), mutta seurantavaiheessa p -arvo on nyt merkitsevyyden rajan toisella puolen ($p = 0.090$).

Tutkimusten pois jättäminen näkyy loppuvaiheen metaregressiossa siten, että kontrolliryhmä ei enää ole merkitsevä kovariaatti ($p = 0.217$). Regressiomalli on

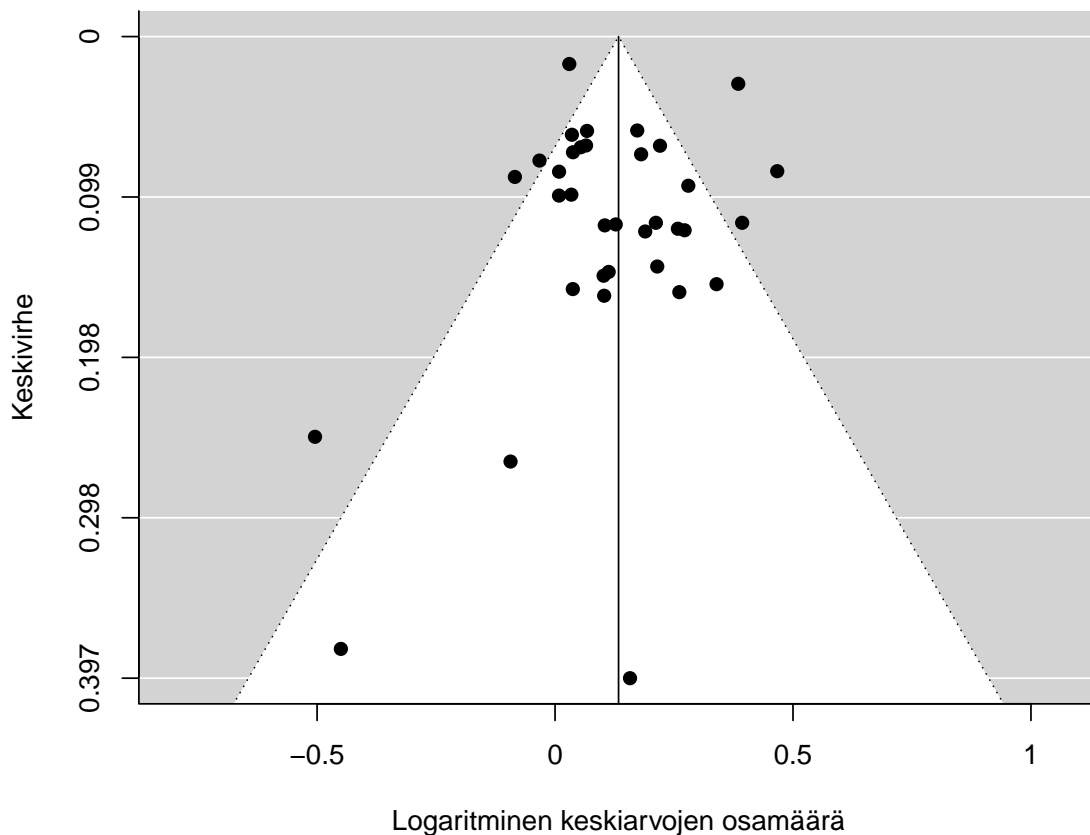
$$\begin{aligned} \hat{R} &= \exp(-0.0351 + 0.0300 \cdot \text{quality}) \\ &\approx 0.966 \cdot 1.030^{\text{quality}}, \end{aligned}$$

jossa laadun kerroin on suunnilleen sama kuin kappalessa 5.1, mutta vaikutusten poistaminen ilmenee suurempana epävarmuutena ($p = 0.036$, luottamusväli [1.001, 1.060]). Heterogeenisyyden selitysaste on 26 %, ja jäännösheterogeenisyys on merkitsevää ($p = 8 \cdot 10^{-8}$). Seurantavaiheessa metaregressiomalliksi saadaan

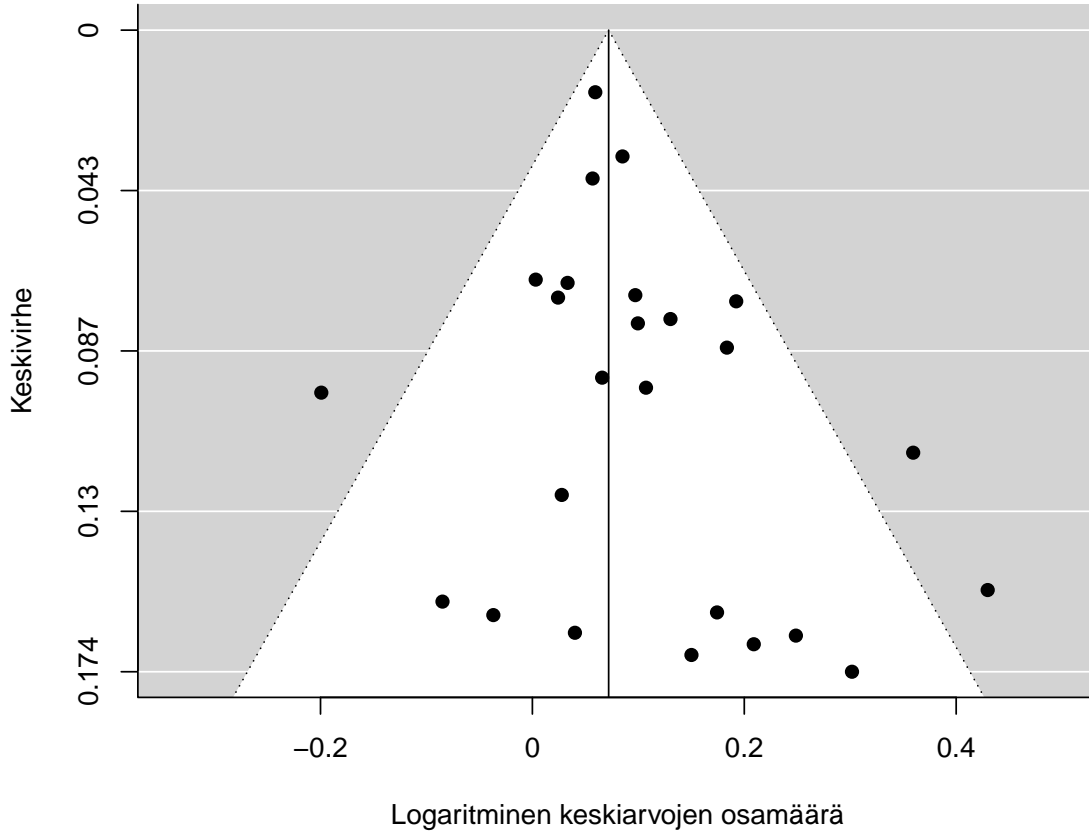
$$\begin{aligned}\hat{R} &= \exp(0.0607 + 0.0622 \cdot multispec) \\ &\approx 1.063 \cdot 1.064^{multispec}.\end{aligned}$$

Mallissa oleva kovariaatti on sama kuin kappalessa 5.3, mutta sen regressiokerroin ja vakiotermit poikkeavat jonkin verran aiemmasta. Yhden asiantuntijan antaman intervention hyöty suhteessa kontrolliryhmän interventioon on noin 6.3 % ($p = 3 \cdot 10^{-6}$, luottamusväli [1.036, 1.090]) ja usean asiantuntijan hyöty noin 13 %. Kovariaatti selittää 16 % heterogeenisyydestä, joka ei tosin ollut alun perinkään merkitsevän suurta eikä ole nytkään ($p = 0.17$)

Keskivirheeltään suurten vaikutusten poistamisen taustalla ollut tavoite, julkaisuharhan eliminoiminen, onnistui vähintään hyvin. Loppuvaiheessa Kendallin $\tau = 0.005$ ($p = 0.978$) ja seurantavaiheessa $\tau = 0.249$ ($p = 0.078$), joten mahdollinen harha on vähäisempää kuin kaikki tutkimukset mukaan otettuna eikä ainakaan merkitsevän suurta. Samaa kielivät hyvin symmetriset suppilokuviot (kuvat 10 ja 11).



Kuva 10: Suppilokuvio loppuvaiheen herkkyysanalyysistä.



Kuva 11: Suppilokuvio seurantavaiheen herkkyysanalyysistä.

5.5 Riippumattomat vaikutukset

Vaikka osa vaikutuksista riippuu toisistaan yhteisen kontrolliryhmän kautta, riippuvuuden sivuuttaminen ei muuta tuloksia nimeksikään. Loppuvaiheessa keskiarvojen osamäärän estimaatti $\hat{R} = 1.176$ ja luottamusväli $[1.106, 1.249]$ ($p = 2 \cdot 10^{-7}$) ovat likipitään samat kuin edellä. Regressiomalliksi valikoituu

$$\hat{R} \approx 1.122 \cdot 1.037^{\text{quality}} \cdot 0.935^{\text{control}},$$

missä kovariaatit regressiokertoimineen ovat melkein identtiset aiempiin nähden; vain vakio poikkeaa edes jotenkin mainittavasti kappaleesta 5.1. Laadun luottamusväli on $[1.009, 1.066]$ ($p = 0.010$) ja kontrolliryhmän tyyppin luottamusväli $[0.886, 0.987]$ ($p = 0.015$) ovat nekin samankaltaiset kuin edellä. Odotetusti nytkin heterogeenisuus on selvästi merkitsevää sekä ilman kovariaatteja ($p = 5 \cdot 10^{-22}$) että niiden kanssa ($p = 6 \cdot 10^{-7}$, selitysaste 43 %).

Seurantavaiheessa ilman Di Loretoa ym. (2003) keskiarvojen osamäärän estimaatti on $\hat{R} = 1.079$, jonka luottamusväli on $[1.055, 1.104]$ ($p = 7 \cdot 10^{-11}$). Heterogeensyyttä koskeva p-arvo on suurempi mutta ei tilastollisesti merkitsevä ($p = 0.033$). Metaregression tuloksena saadaan

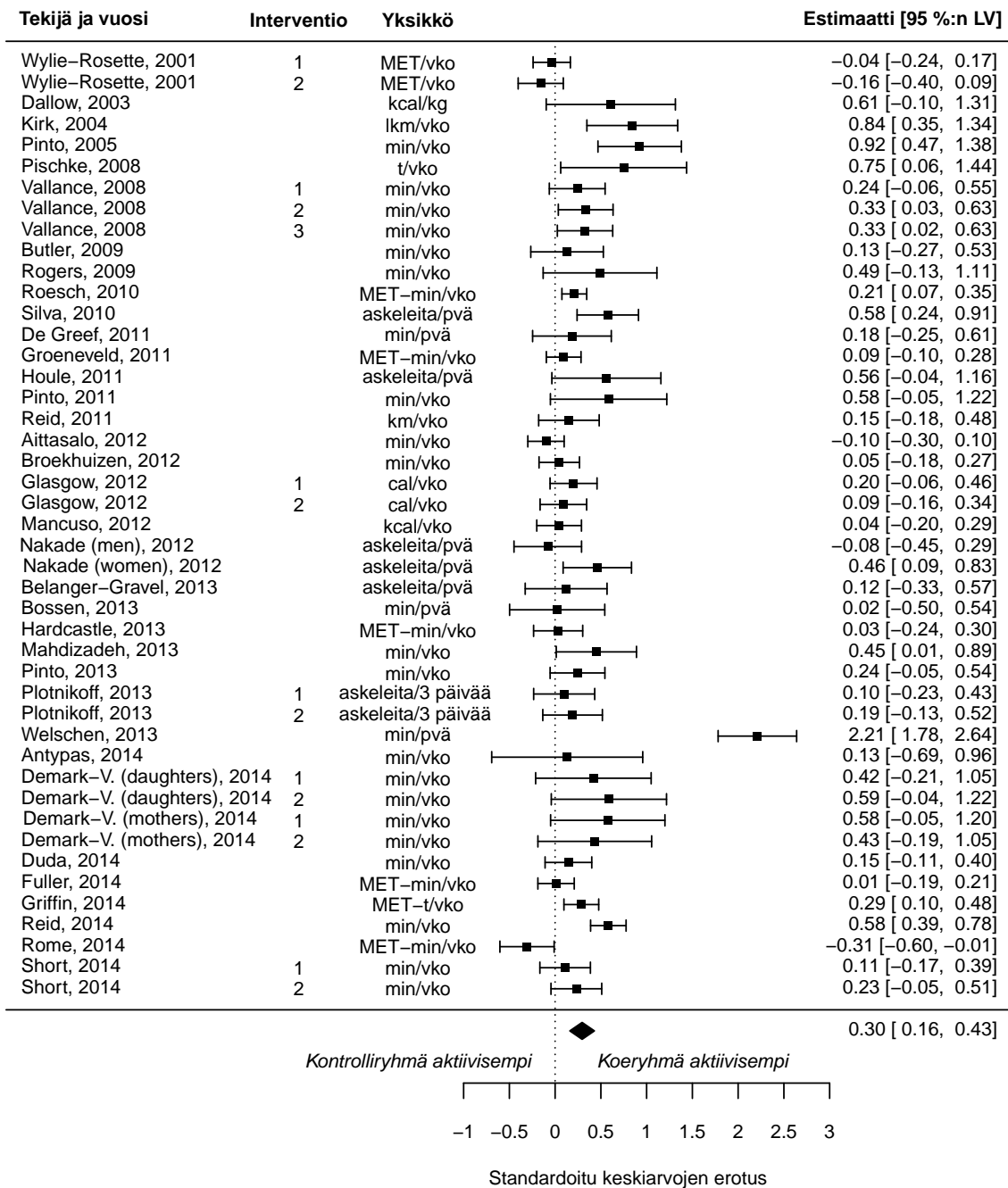
$$\hat{R} \approx 1.065 \cdot 1.070^{\text{multispec}}.$$

Vakiotermin luottamusväli on $[1.039, 1.093]$ ja $p = 10^{-6}$, regressiokertoimella vastaavasti $[1.009, 1.133]$ ja 0.023 . Kertoimet poikkeavat hieman kappaleen 5.3 kertoimista ja niiden luottamusvälit ovat vähän leveämmät, mutta tulkinnan kannalta eroa ei ole. Jäännösheterogeenisuus ei ole nytkään merkitsevän suurta ($p = 0.074$, selitysaste 61 %).

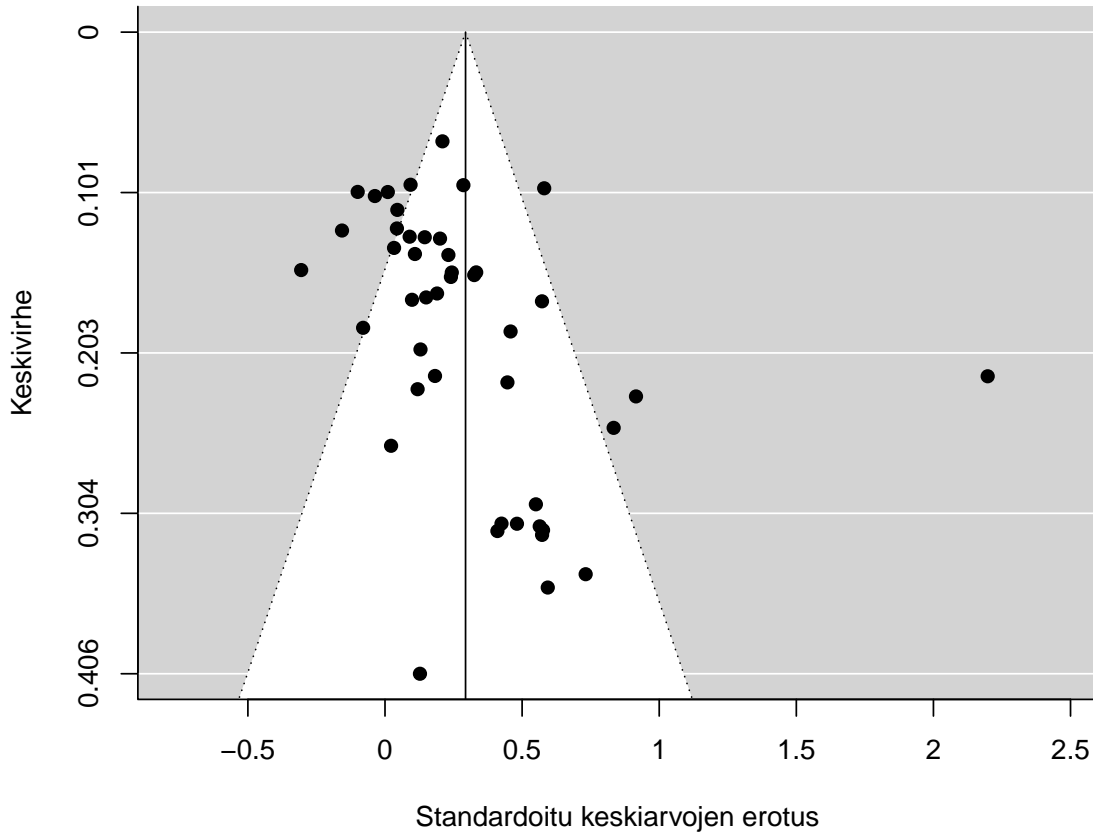
5.6 Standardoitu keskiarvojen erotus

Standardoidulla keskiarvojen erotuksella tulokset ovat samansuuntaisia kuin keskiarvojen osamäärällä, mutta erojakin on. Loppuvaiheessa estimaatti $\hat{d} = 0.295$ on merkitsevästi nollaa suurempi ($p = 2 \cdot 10^{-5}$, 95 %:n luottamusväli $[0.161, 0.429]$, ks. kuva 12), ja vaikutukset ovat heterogeenisia ($p = 2 \cdot 10^{-18}$, ks. kuva 13) Mahdollisesti julkaisuharhasta kertova Kendallin τ on suurempi ja selvästi merkitsevämpi kuin keskiarvojen osamäärällä ($\tau = 0.364, p = 3 \cdot 10^{-4}$).

Regressiomalliksi valikoituu $\hat{d} \approx -0.194 + 0.086 \cdot \textit{quality}$, missä laadun kertoimen p-arvo on 0.024 . Nytkin koeryhmän fyysisellä aktiivisuudella ja tutkimuksen laatu-pisteillä näyttää olevan positiivinen riippuvuus, joskaan kontrolliryhmän ohjeistuksen taso ei ole regressiomallissa mukana. Jos laatu-pisteet vaihtelevat 2–8, on erotus välillä $[-0.194 + 0.086 \cdot 2, -0.194 + 0.086 \cdot 8] \approx [-0.022, 0.494]$. Kuten kappaleessa 5.1 nytkin on suhtauduttava kriittisesti ennusteisiin, joiden mukaan koeryhmä on kontrolliryhmää heikompi. Selitysaste on 21 %, joten tutkimuksen laatu yksistään selittää vain osan heterogeenisyydestä ($p = 2 \cdot 10^{-13}$).



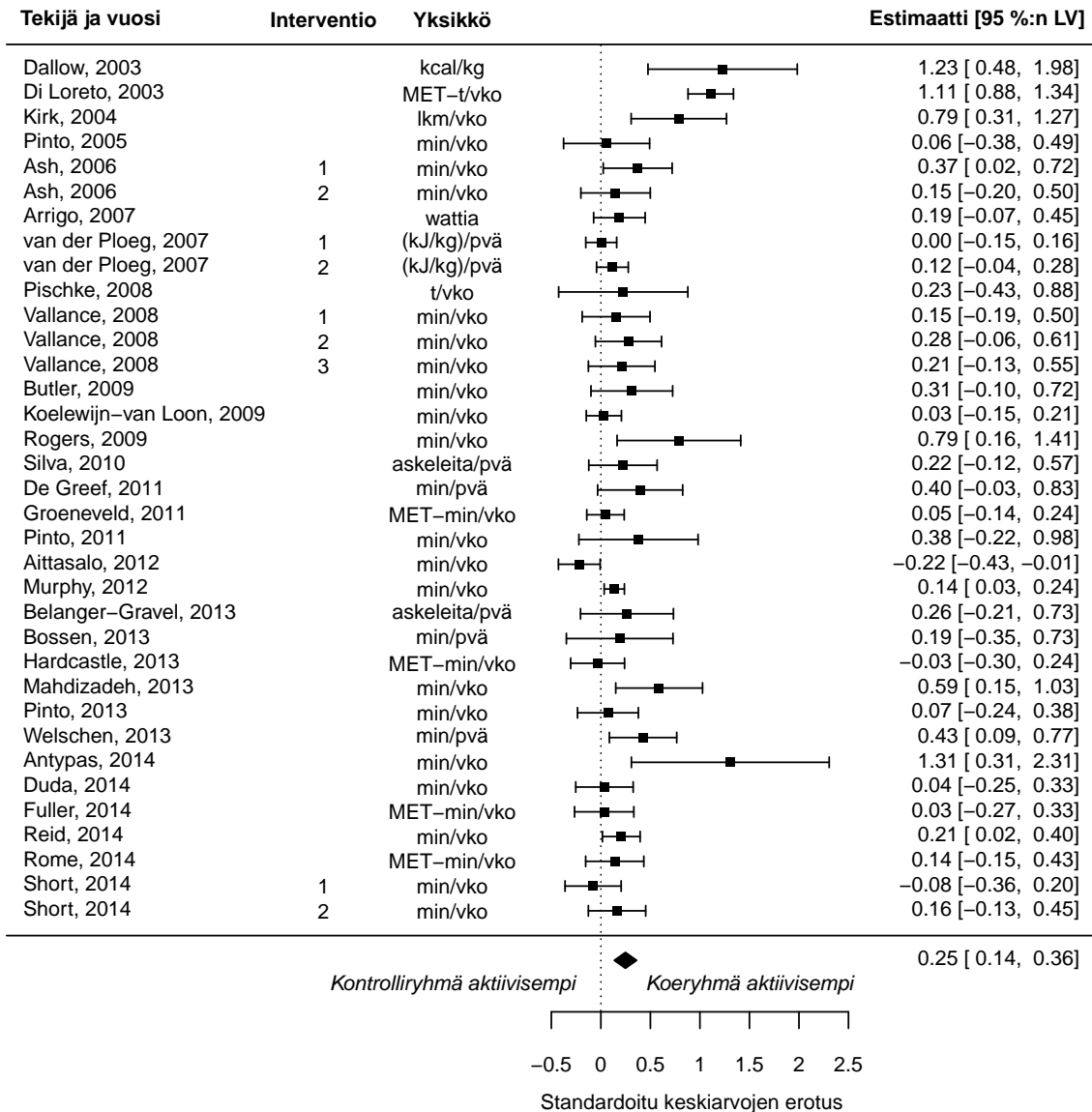
Kuva 12: Metsäkuviio loppuvaiheen meta-analyysistä ilman kovariaatteja, kun vaikutuksena käytetään standardoitua keskiarvojen erotusta.



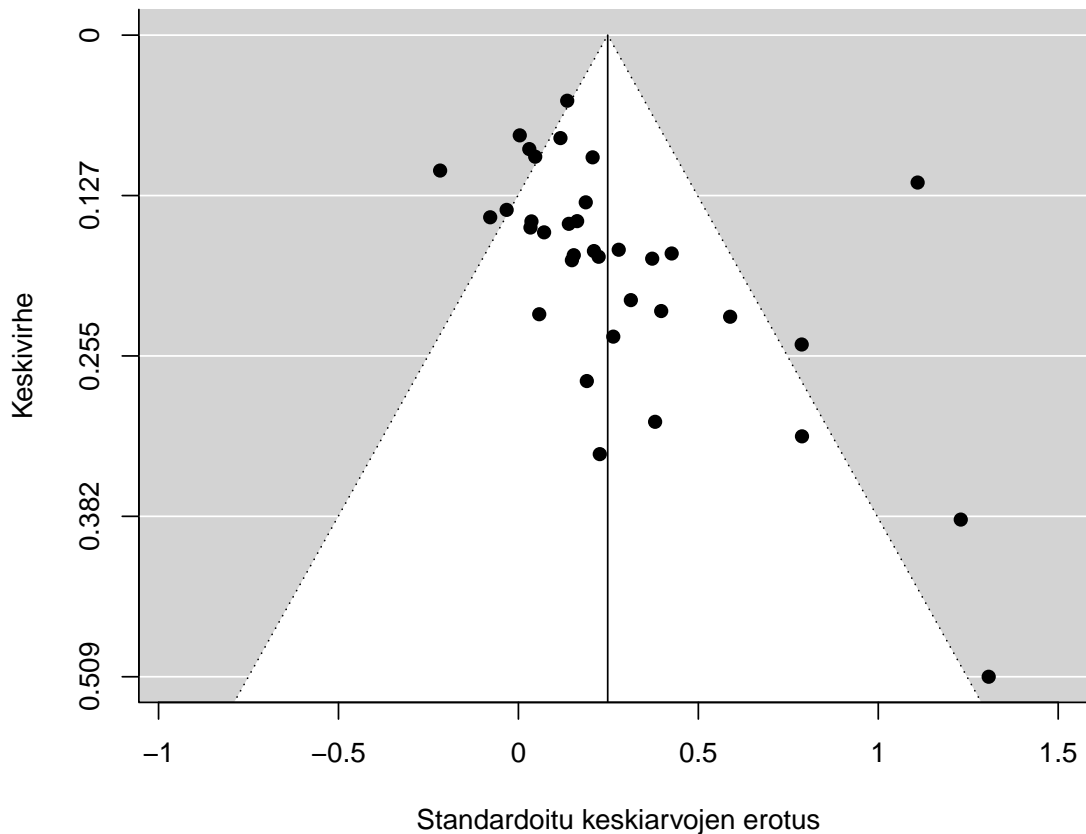
Kuva 13: Suppilokuvio lopputulosten analyysistä ilman kovariaatteja standardoitu keskiarvojen erotusta käytettäessä.

Seurantavaiheessa Di Loreto ym. (2009) mukaan otettuna saadaan $\hat{d} = 0.247$ ($p = 10^{-5}$, 95 %:n luottamusväli [0.136, 0.359]), mikä on pienempi kuin lopputulosten osamäärällä (vrt. esim. kuvia 6 ja 14), jossa Di Loreto ym. (2009) yksin vaikuttaa tuloksiin huomattavasti. Toisaalta on mielenkiintoista huomata, että aiemmin lopputulosten osamäärällä Welschen ym. (2013) näyttäytyy muista poikkeavana havaintona (kuva 12), vaikka näin ei ollut keskiarvojen osamäärää käytettäessä (kuva 5). Havaintojen poikkeavuus voi näin riippua valitusta vaikutuksesta. Vaikutukset ovat merkittävästi heterogeenisiä ($p = 8 \cdot 10^{-13}$). Kendallin τ suurempi kuin keskiarvojen osamäärällä tehdyssä meta-analyysissä ($\tau = 0.496, p = 10^{-5}$), ja kuvassa 15 onkin merkkejä mahdollisesta julkaisuharhasta.

Selkein ero kappaleen 5.2 tuloksiin on metaregressiossa. Malliksi valikoituu $\hat{d} = -0.034 + 0.331 \cdot \mathbb{1}\{\text{health} = 1\} + 0.181 \cdot \mathbb{1}\{\text{health} = 2\}$. Sairaudesta selvinneiden ja terveiden välinen ero ei ole merkittävä ($p = 0.240$), mutta luottamusväli $[-0.121, 0.484]$ on todella leveä. Kohtuullisesta selityksasteesta (32 %) huolimatta heterogeenisyys on selvästi merkittävä ($p = 3 \cdot 10^{-10}$).



Kuva 14: Metsäkuvio seurantavaiheen meta-analyysistä ilman kovariaatteja standardoidun keskiarvojen erotuksen tapauksessa.



Kuva 15: Suppilokuvio seurantavaiheen analyysistä ilman kovariaatteja, kun vaikutus on standardoitu keskiarvojen erotus.

6 Pohdintaa

Esitettyjen tulosten valossa motivointikeinot lisäävät kuntoutuksessa olevien vähän liikkuvien aikuisten fyysistä aktiivisuutta lyhyellä aikavälillä noin 14–17 %, kun vertailukohtana on interventio ilman motivointikeinoja. Fyysisen aktiivisuuden lisäys säilyy osittain intervention jälkeenkin, mutta ei täysin; ero tavanomaiseen interventioon on seurantavaiheen päätteeksi keskimäärin vajaa kahdeksan prosenttia. Seurantavaiheen yhden poikkeavan havainnon (Di Loreto ym. 2003) sisällyttävissä ja poissulkevissa analyyseissä on suuri ero. Kyseinen poikkeava havainto on jokseenkin epäuskottava, eikä sama tutkimus ollut ollenkaan mukana loppuvaiheessa. Pelkästään jo loppuvaiheen ja seurantavaiheen tulosten vertailukelpoisuuden takia tulokset ilman Di Loretoa ym. (2003) ovat tarkoituksenmukaisempia kuin tulokset sen kanssa.

Minkään yksittäisen motivointikeinon suosittelu on hankalaa. Metaregressiossa yksikään motivointikeinoista ei nouse tilastollisesti merkitseväksi kovariaatiksi. Eroja keinojen välillä varmasti on, mutta pienuudessaan ne hukkuvat sattuman sekaan. Toisaalta koska suuria eroja motivointikeinojen välillä ei ole, mitkään niistä eivät ole todistetusti kovin huonoja. Siksi motivointikeinojen käyttöä voidaan yleis-

sesti suosittelalla.

Kovin moni muukaan kovariaatti ei osoittaudu tilastollisesti merkitseväksi. Muutaman kovariaatin kohdalla puuttuva tieto on osasyyp tähän, mutta yleisesti näin ei ole. Loppuvaiheessa tutkimuksen toteutuksen tasoa kuvaavat laatuasteet saattavat olla positiivisesti yhteydessä fyysisen aktiivisuuden eroon koe- ja kontrolliryhmän välillä. Tämä on hiukan yllättävä löytö, sillä pienet laatuasteet voivat kieliä muun muassa valikoivasta raportoinnista ja muista tutkimuksen tekoon liittyvistä rikkomuksista ja edelleen johtaa näennäisen hyviin tuloksiin koeryhmän intervention kannalta. Saatujen tulosten perusteella tilanne on kuitenkin päinvastoin. Voi olla että laadukkaasti raportoidut tutkimukset toteutetaan muutenkin laadukkaasti, mikä näkyy myönteisesti koeryhmän fyysisessä aktiivisuudessa.

Niin ikään loppuvaiheen metaregressiomalliin seuloutunut kontrolliryhmän ohjeistuksen tasoa kuvaava kovariaatti on odotettu; mitä enemmän kontrolliryhmää ohjeistetaan aktiivisuuteen, sitä vaikeampaa koeryhmän on olla aktiivisempi kuin kontrolliryhmä. Toki vastaava ohjeistus kuuluu myös koeryhmän interventioon, mutta joka tapauksessa se rajaa mahdollisuuksia olla kontrolliryhmää aktiivisempi.

Seurantavaiheessa merkitseväksi nousee interventioita antavien asiantuntijoiden lukumäärää kuvaava kovariaatti. Jos asiantuntijoita on vain yksi, ero ryhmien välillä on keskimäärin reilu kuusi prosenttia. Usean asiantuntijan interventioissa ero on noin 13–14 %. Tulos vaikuttaa järkeenkäyvältä; mitä enemmän asiantuntemusta on käytössä, sitä parempia ovat tulokset.

Aineistossa on viitteitä julkaisuharhasta, mutta sen eliminoiminen ei muuta tuloksia kovin paljoa. Meta-analyysin tulokset ilman kovariaatteja pysyvät liki muuttumattomina. Regressiokertoimet sen sijaan joko pienenevät (laatu, asiantuntijoiden lukumäärä) tai muuttuvat jopa merkitsemättömiksi (kontrolliryhmän ohjeistus). Tämä kertonee kuitenkin enemmän regressiomallien ja p-arvojen herkkyydestä kuin merkittävästä julkaisuharhasta. Jos kymmenkunta havaintoyksikköä muutaman kymmenen havaintoyksikön aineistosta poistetaan, tuloksilla on tapana muuttua, vaikka poistetut vaikutukset olisivatkin painoarvoltaan pienimmästä päästä. Regressiokertoimien ja p-arvojen tulkinnessa on syytä olla varovainen varsinkin siksi, että testattuja kovariaatteja on liki kolmekymmentä.

Yhteisistä kontrolliryhmistä juontuva vaikutusten välinen riippuvuus sisällytettiin meta-analyysiin, mutta riippuvuuden sivuuttaminen ei muuta tuloksia nimeksikään. Tämä on ymmärrettävää, sillä sekä loppu- että seurantavaiheessa vain noin prosentti kovariansseista poikkeaa nolasta; suhteellisen harvoilla vaikutuksilla on yhteinen kontrolliryhmä.

Vaihtoehtoisena vaikutuksena kokeiltu standardoitu keskiarvojen erotus tuottaa osin samankaltaisia ja osin erilaisia tuloksia kuin keskiarvojen osamäärä. Ilman kovariaatteja saadaan jälleen merkitsevä ero koeryhmän hyväksi sekä loppu- että seurantavaiheessa. Loppuvaiheen metaregressiossa vain laatu on merkitsevä, joskin tälläkin kertaa sen yhteys fyysiseen aktiivisuuteen koeryhmässä on positiivinen. Seuranta-

vaiheen metaregressiossa eroavuudet ovat suurimmillaan, sillä ainoa merkitsevä kovariaatti on osallistujien terveydentilaa kuvaava kovariaatti; sairast hyötyvät koe-ryhmän interventtiosta terveitä enemmän.

Keskiarvojen osamäärän ja standardoidun keskiarvojen erotuksen laittaminen paremmuusjärjestykseen ei ole yksinkertaista. Selkein ero vaikutusten välillä on tulkin-
nassa, sillä keskiarvojen osamäärä on helppo ymmärtää. Mielenkiintoista on, että eri havainnot ovat poikkeavia eri vaikutuksille; keskiarvojen osamäärä seurantavai-
heessa Di Loreton ym. (2003) tutkimuksessa on moninkertainen muihin nähden, kun taas loppuvaiheessa Welschenin ym. (2013) tutkimus eroaa standardoidulla keskiar-
vojen erotuksellaan muista. Mahdollista julkaisuharhaa mittaava korrelaatiokerroin on syystä tai toisesta järjestelmällisesti suurempi standardoidulla keskiarvojen ero-
tuksella. Ilmiön syytä ei tässä työssä kuitenkaan pyritä selvittämään. Vaikutuksen valinta palautuu lopulta käsitykseen mitattavasta ilmiöstä; ovatko merkityksellisiä erotukset vai logaritmiset erotukset eli osamäärät? Kysymyksen ratkaisun on tultava meta-analyysin ulkopuolelta sovellusalan osajilta.

Meta-analyysin anti systemaattiselle kirjallisuuskatsaukselle on sama kuin yleises-
ti tilastollisten menetelmien numeeristen aineistojen analysoinnille: teoriaan perus-
tuva yhteenvedo aineistosta pelkän kuvailun ja subjektiivisten näkemysten sijaan. Meta-analyysillä voidaan painottaa havaintoyksiköitä niiden koon mukaan; tuhan-
nen osallistujan tutkimus on tärkeämpi kuin parinkymmenen osallistujan. Mitatta-
vaa ominaisuutta voidaan mallintaa taustamuuttujien ehdoilla hyödyntämällä me-
taregressiota.

Myös meta-analyysin ongelmat ja rajoitteet ovat tilastotieteelle tyypillisiä. Hypo-
teesien testaus ja mallinvalinta ovat haastavia tehtäviä, jos kovariaatteja on paljon. Toisaalta jokin ilmiön kannalta olennainen kovariaatti voi puuttua, jos tutkimusar-
tikkeleissa ei ole sitä raportoitu tai sen mahdollista yhteyttä tutkittavaan ilmiöön ei ole tunnistettu. Vaikka tavoitteena on objektiivinen vastaus tutkimuskysymyk-
siin, on menetelmällisten yksityiskohtien, kuten vaikutuksen, valinnalla merkitystä johtopäätöksiä tehtäessä.

Lähteet

Ainsworth, B. E. – Haskell, W. L. – Whitt, M. C. – Irwin, M. L. – Swartz, A. M. – Strath, S. J. – O’Brien, W. L. – Basset Jr., D. R. – Schmitz, K. H. – Emplaincourt, P. O. – Jacobs Jr., D. R. – Leon, A. S. 2000: Compendium of Physical Activities: an update of activity codes and MET intensities. *Medicine & Science in Sports & Exercise*, 32(supplement), S498–S516.

Aittasalo, M. – Raitanen, J. – Kinnunen, T. I. – Ojala, K. – Kolu P. – Luoto, R. 2012: Is intensive counseling in maternity care feasible and effective in promoting physical activity among women at risk for gestational diabetes? Secondary analysis of a cluster randomized NELLI study in Finland. *International Journal of Behavioral Nutrition and Physical Activity*, 9(104).

Antypas, K. – Wangberg, S. C. 2014: An Internet- and Mobile-Based Tailored Intervention to Enhance Maintenance of Physical Activity After Cardiac Rehabilitation: Short-Term Results of a Randomized Controlled Trial. *Journal of Medical Internet Research*, 16(3).

Arrigo, I. – Brunner-LaRocca H. – Lefkovits, M. – Pfisterer, M. – Hoffmann, A. 2008: Comparative outcome one year after formal cardiac rehabilitation: the effects of a randomized intervention to improve exercise adherence. *European Journal of Cardiovascular Prevention and Rehabilitation*, 15(3), 306–311.

Ash, S. – Reeves, M. – Bauer, J. – Dover, T. – Vivanti, A. – Leong, C. – O’Moore Sullivan, T. 2006: A randomised control trial comparing lifestyle groups, individual counselling and written information in the management of weight and health outcomes over 12 months. *International Journal of Obesity*, 30(10), 1557–1564.

Bauman, A. 2004: Updating the evidence that physical activity is good for health – an epidemiological review 2000–2003. *Journal of Science and Medicine in Sport*, 7(1) Supplement, 6–19.

Begg, C. B. – Mazumdar, M. 1994: Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*, 50(4), 1088–1101.

Bèlanger-Gravel, A. – Godin, G. – Bilodeau, A. – Poirier, P. 2013: The effect of implementation intentions on physical activity among obese older adults: A randomized control study. *Psychology & Health*, 28(2), 217–233.

Borenstein, M. – Hedges, L. V., – Higgins, J. P. T. – Rothstein, H. R. 2009: *Introduction to Meta-analysis*. Wiley, Hoboken.

Bossen, D. – Veenhof, C. – Van Beek, K. EC. – Spreeuwenberg P. MM. – Dekker, J. – De Bakker, D. H. 2013: Effectiveness of a Web-Based Physical Activity Inter-

vention in Patients With Knee and/or Hip Osteoarthritis: Randomized Controlled Trial. *Journal of Medical Internet Research*, 15(11).

Broekhuizen, K. – van Poppel, M. N. M. – Koppes, L. L. – Kindt, I. – Brug, J. – van Mechelen, W. 2012: Can Multiple Lifestyle Behaviours Be Improved in People with Familial Hypercholesterolemia? Results of a Parallel Randomised Controlled Trial. *PLoS ONE*, 7(12).

Butler, L. – Furber, S. – Phongsavan, P. – Mark, A. – Bauman, A. 2009: Effects of a Pedometer-Based Intervention on Physical Activity Levels After Cardiac Rehabilitation: A RANDOMIZED CONTROLLED TRIAL. *Journal of Cardiopulmonary Rehabilitation and Prevention*, 29(2), 105–114.

Chasandra, M. – Novoradovskaya, E. – Sjögren, T. – Lintunen, T. – Karvanen J. – Immonen, J. – Ruiz, M. 2017: Effectiveness of motivational methods to change physical activity behavior of sedentary adults in rehabilitation settings: A systematic review with meta-analyses. *Käsikirjoitus*.

Dallow, C. B. – Anderson, J. 2003: Using Self-efficacy and a Transtheoretical Model to Develop a Physical Activity Intervention for Obese Women. *American Journal of Health Promotion*, 17(6), 373–381.

De Greef, K. P. – Deforche, B. I. – Ruige, J. B. – Bouckaert, J. J. – Tudor-Locke, C. E. – Kaufman, J-M. – De Bourdeaudhuij, I. M. 2011: The effects of a pedometer-based behavioral modification program with telephone support on physical activity and sedentary behavior in type 2 diabetes patients. *Patient Education and Counseling*, 84(2), 275–279.

Demark-Vahnefried, W. – Jones, L. W. – Snyder, D. C. – Sloane, R. J. – Kimmick, G. G. – Hughes, D. – Badr, H. J. – Miller, P. E. – Burke, L. E. – Lipkus, I. M. 2014: Daughters and Mothers Against Breast Cancer (DAMES): Main Outcomes of a Randomized Controlled Trial of Weight Loss in Overweight Mothers With Breast Cancer and Their Overweight Daughters. *Cancer*, 120(16), 2522–2534.

Di Loreto, C. – Parlanti, N. – Fanelli, C. – Santeusanio, F. – Lucidi, P. – Brunetti, P. – Murdolo, G. – De Feo, P. – De Cicco, A. 2003: Validation of a Counseling Strategy to Promote the Adoption and the Maintenance of Physical Activity by Type 2 Diabetic Subjects. *Diabetes Care*, 26(2), 404–408.

Duda, J. – Williams, G. – Ntoumanis, N. – Daley, A. – Eves, F. – Mutrie, N. – Rouse, P. – Lodhia, R. – Blamey, R. – Jolly, K. 2014: Effects of a standard provision versus an autonomy supportive exercise referral programme on physical activity, quality of life and well-being indicators: a cluster randomised controlled trial. *International Journal of Behavioral Nutrition and Physical Activity*, 11(10).

Egger, M. – Smith, G. D. – Schneider, M. – Minder, C. 1997: Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629–634.

Fuller, N. R. – Williams, K. – Shrestha, R. – Ahern, A. L. – Holzapfel, C. – Hauner, H. – Jebb, S. A. – Caterson, I. D. 2014: Changes in physical activity during a weight loss intervention and follow-up: a randomized controlled trial. *Clinical Obesity*, 4(3), 127–135.

Gilbert, R. – Salanti, G. – Harden, M. – See, S. 2005: Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *International Journal of Epidemiology*, 34, 874–887.

Glasgow, R. E. – Kurz, D. – King, D. – Dickman, J. M. – Faber, A. J. – Halterman, E. – Woolley, T. – Toobert, D. J. – Strycker, L. A. – Estabrooks, P. A. – Osuna, D. – Ritzwoller, D. 2012: Twelve-Month Outcomes of an Internet-Based Diabetes Self-Management Support Program. *Patient Education and Counseling*, 87(1), 81–92.

Gleser, L. J. – Olkin, I. 2009: Stochastically Dependent Effect Sizes. Teoksessa Cooper, H. M. – Hedges, L. V. – Valentine, J.C. (toim.): *The Handbook of Research Synthesis and Meta-analysis*, 357–376. Sage, New York.

Griffin, S. J. – Simmons, R. K. – Prevost, A. T. – Williams, K. M. – Hardeman, W. – Sutton, S. – Brage, S. – Ekelund, U. – Parker, R. A. – Wareham, N. J. – Kinmonth, A. L. 2014: Multiple behaviour change intervention and outcomes in recently diagnosed type 2 diabetes: the ADDITION-Plus randomised controlled trial. *Diabetologia*, 57(7), 1308–1219.

Groeneveld, I. – Proper, K. – van der Beek, A. J. – Hildebrandt, V. H. – van Mechelen, W. 2011: Short and long term effects of a lifestyle intervention for construction workers at risk for cardiovascular disease: a randomized controlled trial. *BioMed Central Public Health*, 11(836).

Hardcastle, S. J. – Taylor, A. H. – Bailey M. P. – Harley, R. A. – Hagger, M. S. 2013: Effectiveness of a motivational interviewing intervention on weight loss, physical activity and cardiovascular disease risk factors: a randomised controlled trial with a 12-month post-intervention follow-up. *International Journal of Behavioral Nutrition and Physical Activity*, 10(40).

Hedges, L.V. – Gurevitch, J. – Curtis, P. S. 1999: The Meta-analysis of Response Ratios in Experimental Ecology. *Ecology* 80(4), 1150–1156.

Hedges, L. V. – Olkin, I. 1985: *Statistical Methods for Meta-analysis*. Academic Press, San Diego.

- Hedges, L. V. – Vevea, J. L. 1998: Fixed- and Random-effects Models in Meta-analysis. *Psychological Methods* 3(4), 486–504.
- Ver Hoef, J. M. 2012: Who Invented the Delta Method? *The American Statistician* 66(2), 124–127.
- Houle, J. – Doyon, O. – Vadeboncoeur, N. – Turbide, G. – Diaz, A. – Poirier, P. 2011: Innovative program to increase physical activity following an acute coronary syndrome: Randomized controlled trial. *Patient Education and Counseling* 85, e237–e244.
- Knapp, G. – Hartung, J. 2003: Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17), 2693–2710.
- Kirk, A. F. – Mutrie, N. – MacIntyre, P. D. – Fisher, M. B. 2004: Promoting and Maintaining Physical Activity in People with Type 2 Diabetes. *American Journal of Preventive Medicine*, 27(4), 289–296.
- Koelewijn-van Loon M. S. – van der Weijden, T. – van Steenkiste, B. – Ronda, G. – Winkens, B. – Severens, J. L. – Wensing, M. – Elwyn, G. – Grol, R. 2009: Involving patients in cardiovascular risk management with nurse-led clinics: a cluster randomized controlled trial. *Canadian Medical Association. Journal*, 181(12).
- Konstantopoulos, S. 2011: Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, 2(1), 61–76.
- Lajeunesse, M. J. 2011: On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology*, 92(11), 2049–2055.
- Mahdizadeh, M. – Peymam, N. – Taghibour, A. – Esmaily, H. – Mahdizadeh, S. M. 2013: Effect of Health Education Program on Promoting Physical Activity among Diabetic Women in Mashhad, Iran: Applying Social Cognitive Theory. *Journal of Research in Health Sciences*, 13(1), 90–97.
- Mancuso, C. A. – Choi, T. N. – Westermann, H. – Wenderoth, S. – Hollenberg, J. P. – Wells, M. T. – Isen, A. M. – Jobe, J. B. – Allegrante, J. P. – Charlson, M. E. 2012: Increasing Physical Activity in Patients with Asthma Through Positive Affect and Self-affirmation: A Randomized Trial. *Archives of Internal Medicine*, 172(4), 337–343.
- Murphy, S. M. – Edwards, R. T. – Williams, N. – Raisanen, L. – Moore, G. – Linck, P. – Hounsome, N. – Din, N. U. – Moore, L. 2012: An evaluation of the effectiveness and cost effectiveness of the National Exercise Referral Scheme in Wales, UK: a randomised controlled trial of a public health policy initiative. *Journal of Epidemiological Community Health*, 66, 745–753.

- Nakade, M. – Aiba, N. – Suda, N. – Morita, A. – Motohiko, M. – Sasaki, S. – Watanabe, S. 2012: Behavioral change during weight loss program and one-year follow-up: Saku Control Obesity Program (SCOP) in Japan. *Asia Pacific Journal of Clinical Nutrition*, 21(1), 22–34.
- Pinto, B. M. – Frierson, G. M. – Rabin, C. – Trunzo, J. J. – Marcus, B. H. 2005: Home-Based Physical Activity Intervention for Breast Cancer Patients. *Journal of Clinical Oncology*, 23(15), 3577–3587.
- Pinto, B. M. – Papandonatos, G. D. – Goldstein, M. G. 2013: Randomized Trial to Promote Physical Activity Among Breast Cancer Patients. *Health Psychology*, 32(6), 616–626.
- Pinto, B. M. – Papandonatos, G. D. – Goldstein, M. G. – Marcus B. H. – Farrell, N. 2013: Home-based physical activity intervention for colorectal cancer survivors. *Psycho-Oncology*, 22(1), 54–64.
- Pischke, C. R. – Scherwitz, L. – Weidner, G. – Ornish, D. 2008: Long-Term Effects of Lifestyle Changes on Well-Being and Cardiac Variables Among Coronary Heart Disease Patients. *Health Psychology*, 27(5), 584–592.
- Plotnikoff, R. C. – Karunamuni, N. – Courneya, K. S. – Sigal, R. J. – Johnson J. A. – Johnson, S. T. 2013: The Alberta Diabetes and Physical Activity Trial (ADAPT): A Randomized Trial Evaluating Theory-Based Interventions to Increase Physical Activity in Adults with Type 2 Diabetes. *Annals of Behavioral Medicine*, 45, 45–56.
- Raudenbush, S. W. 2009: Analyzing Effect Sizes: Random-effects models. Teoksessa Cooper, H. M. – Hedges, L. V. – Valentine, J.C. (toim.): *The Handbook of Research Synthesis and Meta-analysis*, 295–315. Sage, New York.
- R Core Team 2017: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Wien. URL <https://www.R-project.org/> (viitattu 13.4.2017).
- Reid, R. D. – McDonnell, L. A. – Riley, D. L. – Mark, A. E. – Mosca, L. – Beaton, L. – Papadakis, S. – Blanchard, C. M. – Mochari-Greenberger, H. – O’Farrell, P. – Wells, G. A. – Slovinec D’Angelo, M. E. – Pipe, A. L. 2014: Effect of an intervention to improve the cardiovascular health of family members of patients with coronary artery disease: a randomized trial. *Canadian Medical Association Journal*, 186(1), 23–30.
- Reid, R. D. – Morrin, L. I. – Higginson, L. A.J. – Wielgosz, A. – Blanchard, C. – Beaton, L. J. – Nelson, C. – McDonnell, L. – Oldridge, N. – Wells, G. A. – Pipe, A. L. 2011: Motivational counselling for physical activity in patients with coronary artery disease not participating in cardiac rehabilitation. *European Journal of Car-*

diovascular Prevention & Rehabilitation, 19(2), 161–166.

Roesch, S. – Norman, G. J. – Villodas, F. – Sallis, J. F. – Patrick, K. 2010: Intervention-mediated effects for adult physical activity: A latent growth curve analysis. *Social Science & Medicine*, 71(3), 494–501.

Rogers, L. Q. – Hopkins-Price, P. – Vicari, S. – Markwell, S. – Pamentier, R. – Courneya, K. S. – Hoelzer, K. – Naritoku, C. – Edson, B. – Jones, L. – Dunnington, G. – Verhulst, S. 2009: Physical Activity and Health Outcomes Three Months After Completing a Physical Activity Behavior Change Intervention: Persistent and Delayed Effects. *Cancer, Epidemiology, Biomarkers & Prevention*, 18(5), 1410–1418.

Romé, Å. – Persson, U. – Ekdahl, C. – Gard, G. 2014: Costs and outcomes of an exercise referral programme – A 1-year follow-up study. *European Journal of Physiotherapy*, 16(2), 82–92.

Short, C. E. – James, E. L. – Girgis, A. – D’Souza, M. I. – Plotnikoff, R. C. 2015: Main outcomes of the Move More for Life Trial: a randomised controlled trial examining the effects of tailored print and targeted-print materials for promoting physical activity among post-treatment breast cancer survivors. *Psycho-Oncology*, 24(7), 771–778.

Silva, M. N. – Vieira, P. N. – Coutinho, S. R. – Minderico, C. S. – Matos, M. G. – Sardinha, L. B. – Teixeira, P. J. 2010: Using self-determination theory to promote physical activity and weight control: a randomized controlled trial in women. *Journal of Behavioral Medicine*, 33(2), 110–122.

Sutton, A. J. 2009: Publication bias. Teoksessa Cooper, H. M. – Hedges, L. V. – Valentine, J.C. (toim.): *The Handbook of Research Synthesis and Meta-analysis*, 435–452. Sage, New York.

Vallance, J. K. – Courneya, K. S. – Taylor, L. M. – Plotnikoff, R. C. – Mackey, J. R. 2008: Development and Evaluation of a Theory-Based Physical Activity Guidebook for Breast Cancer Survivors. *Health Education & Behavior*, 35(2), 174–189.

van der Ploeg, H. P. – Streppel, K. R. M. – van der Beek, A. J. – van der Woude, L. H. V. – Vollenbroek-Hutten, M. M. R. – van Harten, W. H. – van Mechelen, W. 2007: Successfully Improving Physical Activity Behavior After Rehabilitation. *American Journal of Health Promotion*, 21(3), 153–159.

Viechtbauer, W. 2010: Conducting Meta-analyses in R with the metafor Package. *Journal of Statistical Software*, 36(3).

Welschen, L. M. C. – van Oppen, P. – Bot, S. D. M. – Kostense, P. J. – Dekker, J. M. – Nijpels, G. 2013: Effects of a cognitive behavioural treatment in patients with

type 2 diabetes when added to managed care; a randomised controlled trial. *Journal of Behavioral Medicine*, 36(6), 556–566.

Wylie-Rosette, J. – Swencionis, C. – Ginsberg, M. – Cimino, C. – Wassertheil-Smoller, S. – Caban, A. – Segal-Isaacson, C.J. – Martin, T. Lewis, J. 2001: Computerized weight loss intervention optimizes staff time: the clinical and cost results of a controlled clinical trial conducted in a managed care setting. *Journal of the American Dietetic Association*, 101(10), 1155–1162.

Liitteet

Liite 1: Loppuvaiheen meta-analyysin R-koodin osa, joka vastaa kappaleen 5.1 tuloksia.

```
options(stringsAsFactors=F) # Metsäkuvioiden piirtämistä varten

moti<-read.table("motivating_final.txt",header=T) # Aineisto

# install.packages("metafor") # metafor-paketti
library(metafor)

# Järjestetään aineisto tekijän ja vuoden mukaan.
# Tästä on hyötyä kovarianssimatriisin muodostamisessa.
q<-moti[order(moti$Tekijä),]

attach(q) # Muuttujat käyttöön

# Lyhennysmerkinnät otoskooille, keskiarvoille ja keskihajonnoille
ni<-finaln_i
nc<-finaln_c
mi<-finalmean_i
mc<-finalmean_c
sdi<-finalsd_i
sdc<-finalsd_c

R_final<-mi/mc # Keskiarvojen osamäärät R

lnR_final<-log(R_final) # Logaritmiset R:t

# Logaritmisten R:ien varianssit
V_lnR_final<-sdc^2/(nc*mc^2)+sdi^2/(ni*mi^2)

# Asetetaan varianssit aineiston viimeiseksi sarakkeeksi.
q[,59]<-V_lnR_final

# Funktio kovarianssimatriisin muodostamiseen
calc.v<-function(x)
  v<-matrix((x$finalsd_c[1])^2/((x$finaln_c[1])*(x$finalmean_c[1])^2),
    nrow=nrow(x),ncol=nrow(x))
  diag(v)<-x[,59]
  v
```

```

# Kovarianssimatriisi
V_final<-bldiag(lapply(split(q,q$Tekijä),calc.v))

# Määritellään erikseen Roesch ym. (2010) ja Broekhuizen ym.(2012),
# sillä niiden tulokset ovat valmiiksi logaritmiasteikolla.
r<-which(Tekijä=="Roesch_2010")
b<-which(Tekijä=="Broekhuizen_2012")

lnR_final[r]<-mi[r]-mc[r]
V_final[r,r]<-sdi[r]^2/ni[r]+sdc[r]^2/nc[r]

lnR_final[b]<-mi[b]-mc[b]
V_final[b,b]<-sdi[b]^2/ni[b]+sdc[b]^2/nc[b]

# Malli ilman kovariaatteja
id<-1:length(lnR)
res_final<-rma.mv(lnR_final,V_final,
method="ML",random = ~ 1 | Tekijä_re/id)
res_final

# Järjestyskorrelaatiotesti
ranktest(res_final)

# Vektori, jolla erotellaan saman tutkimuksen eri vertailut toisistaan
t<-c(rep("_",8),1,2,1,2,"_","_",1,2,rep("_",13),1,2,rep("_",5),1,2,
"_",1,2,3,"_",1,2)

# Metsäkuvion piirtäminen
# Tallennetaan kuva pdf-komennolla pdf-muotoon.
pdf("E:/Meta-analyysi/final_fin.pdf",width=7,height=9)

forest(res_final,refline=1,transf=exp,
slab=paste(q$Tekijä1,q$Vuosi,sep=" "),
ilab=cbind(iffelse(q$Tyyppi=="_", "",t),q$Yksikkö),
ilab.xpos=c(-2.2,-0.7),cex=0.75,lwd=1,
alim=c(0,5),
xlim=c(-6,7.5),
steps=6,
order=order(q$Vuosi,q$Tekijä1),psize=1)

# Otsikot kuvion tiedoille
op<-par(cex=0.75,font=2)
text(-4.65,47,"Main Author and Year")
text(6.4,47,"Estimate [95% CI]")

```

```

text(-2.2,47,"Intervention")
text(-0.7,47,"Unit")

op1<-par(cex=0.75,font=1)
text(0,-1.6,"Favours Control")
text(3,-1.6,"Favours Intervention")

dev.off()

# Suppilokuvio
pdf("E:/Meta-analyysi/funnel_final_fin.pdf",width=7,height=6)

funnel(res_final,xlab="Logaritminen keskiarvojen osamäärä",
ylab="Keskivirhe")

dev.off()

# Kovariaattien testaus

# Tavoitteet ja suunnittelu
goals1<-ifelse(goals==0,0,1)
rma.mv(lnR_final,V_final,mods=~goals1,method="ML",
random = ~ 1 | Tekijä_re/id)

# Palaute ja tarkkailu
feedback1<-ifelse(feedback==0,0,1)
rma.mv(lnR_final,V_final,mods=~feedback1,method="ML",
random = ~ 1 | Tekijä_re/id)

:

# Yksi tai useampi motivointikeino
rma.mv(lnR_final,V_final,mods=~bct,method="ML",
random = ~ 1 | Tekijä_re/id)

# Motivointikeinojen lukumäärä
nbct<-goals+feedback+support+shaping+natural+
comparison_beh+associations+repetition+comparison_out+
reward+regulation+antecedents+identity+scheduled+
selfbelief+covert+increase
rma.mv(lnR_final,V_final,mods=~nbct,method="ML",
random = ~ 1 | Tekijä_re/id)

:

```

```

# Interventiovaiheen kesto kuukausiksi muunnettuna
length_inter_month<-length_inter/4
rma.mv(lnR_final,V_final,mods=~length_inter_month,method="ML",
random = ~ 1 | Tekijä_re/id)

# Ohjeistus luokiteltuna ja jatkuvana
rma.mv(lnR_final,V_final,mods=~factor(control),method="ML",
random = ~ 1 | Tekijä_re/id)
rma.mv(lnR_final,V_final,mods=~control,method="ML",
random = ~ 1 | Tekijä_re/id)

# Mallinvalinnan jälkimmäinen vaihe
malli1<-rma.mv(lnR_final,V_final,mods=~laatu,method="ML",
random = ~ 1 | Tekijä_re/id)
malli2<-rma.mv(lnR_final,V_final,mods=~laatu+control,method="ML",
random = ~ 1 | Tekijä_re/id)

anova(malli1,malli2)

# Lopullinen metaregressiomalli
malli2

# Metaregressiomallin selitysaste
(1-sum(malli2$sigma2)/sum(res_final$sigma2))*100

```