

Vaikuttavuustutkimus:

Mikä on sopivin analyysimenetelmä toistettujen mittausten analysointiin satunnaistetussa vertailukokeessa?

Artikkeli on jatkoa Kasvun tuki -aikakauslehdessä julkaistulle Vaikuttavuustutkimus-artikkelisarjalle:

- Vaikuttavuustutkimus: [Satunnaistettu vertailukoe tulee suunnitella huolella \(2/2021\)](#)
- Vaikuttavuustutkimus: [Satunnaistetun vertailukokeen laadukkaan raportoinnin edellytykset \(1/2022\)](#)
- Vaikuttavuustutkimus: [Miten arvioida harhaa satunnaistetuissa vertailukokeissa? \(2/2022\)](#)
- Vaikuttavuustutkimus: [Mitä tulee huomioida analysoitaessa satunnaistetun vertailukokeen aineistoa? \(1/2023\)](#)
- Vaikuttavuustutkimus: [Meta-analyysistä lisää voimaa tulosten tulkinnan tueksi \(2/2023\)](#)

- Toistomittauksilla tarkoitetaan samalta tutkimukseen osallistuvilta mitattuja toistuvia vastemuuttujien arvoja.
- Useimmiten kiinnostuksen kohteena on tutkia, tapahtuuko ryhmien välillä tilastollisesti merkitsevää muutosta yli ajan eli ajan kuluessa.
- Vaikutusten arvioimiseen käytetään yleisesti joko toistomittausten kovarianssianalyysia, toistomittausten analyysia tai muutosten analyysia.
- Tilastollisen voiman vuoksi suositeltavin tapa on toistomittausten kovarianssianalyysi.
- Toistomittausten analyysin tuloksia on helppo havainnollistaa mallin mukaisilla keskiarvokäyrillä.
- Satunnaistetussa vertailukokeessa vastemuuttujien lähtötasomittaukset tehdään aina ennen satunnaistamista.



**SANNA
HINKKA-YLI-SALOMÄKI**
johtaja, biostatistiikka ja
tiedonhallinta
VTL, MSc, DLSHTM
Lastenpsykiatrian tutkimuskeskus
Turun yliopisto

Satunnaistettu vertailukoe alkaa suunnittelusta. Asianmukaisesti suunniteltua ja laadukkaasti toteutettua satunnaistettua vertailukoetta pidetään yleisesti tieteellisesti laadukkaimpana menetelmänä selvittää eri interventioiden toimivuutta ja mitata niiden vaikuttavuutta (Hinkka-Yli-Salomäki, 2021). Jokaisessa satunnaistetussa vertailukokeessa tutkijoita ohjaa tutkimusryhmän yhdessä laatima tutkimussuunnitelma. Suunni-

telma toimii perustana tutkimuksen rekisteröinnille ja auttaa niin osallistujia kuin ulkopuolisia arvioijia, kuten tutkimuseettisiä toimikuntia, instituutioiden arviointilautakuntia ja rahoittajia, tutkimuksen arvioinnissa. On havaittu, että monet tutkimussuunnitelmat eivät ole riittävän yksityiskohtaisia tai sisällä kaikkia oleellisia tutkimuksen perusasioita (Tunn ym., 2024). Puutteellinen tutkimussuunnitelma heikentää tutkimuksen ymmärtämistä ja toteutusta, vähentää suunnitelman arvioinnin tehokkuutta ja sitä kautta vaikuttaa tutkimuksen laatuun heikentävästi sekä johtaa tutkimusryhmää ja muita toimijoita, kuten eettisiä toimikuntia, kuormittaviin tutkimussuunnitelmien muutosprosesseihin.

”Satunnaistettu vertailukoe alkaa suunnittelusta.”

Keskeisenä tavoitteena on suunnitella jokainen tutkimus siten, että se tuottaa mahdollisimman harhatonta tietoa interventioiden vaikuttavuudesta. Tällöin voidaan arvioida niiden suoraa kansanterveydellistä merkitystä, kliinistä merkityksellisyyttä ja tilastollista tarkkuutta. On olennaisen tärkeää erottaa tilastollisesti merkitsevä ero kliinisesti merkitsevästä erosta. Hyvin laajoissa

tutkimuksissa voi joskus ilmetä tuloksissa pieniä, tilastollisesti merkitseviä eroja, joilla ei kuitenkaan ole kliinistä merkitystä. On myös tärkeää, että tutkimus voidaan viedä onnistuneesti päätökseen. Toisin sanoen tutkimussuunnitelman on oltava toteutettavissa eettisestä, tieteellisestä, organisatorisesta ja rahoituksellisesta näkökulmasta (Pocock ym., 2015).

Tutkimussuunnitelma ja tilastollinen analyysisuunnitelma

Miten teen riittävän yksityiskohtaisen tutkimussuunnitelman? SPIRIT-lausuma (*Standard Protocol Items: Recommendations for Interventional Trials*), joka julkaistiin vuonna 2013, kehitettiin parantamaan tutkimussuunnitelmien laatua (Chan ym., 2013). Se sisältää asioita, joita satunnaistettujen vertailukokeiden tutkimussuunnitelmissa olisi hyvä olla. Näitä ovat esimerkiksi

- tutkimuksen tarkoitus ja hypoteesit
- koasetelma
- osallistujien mukaanotokriteerit
- päävastemuuttajat ja toissijaiset vastemuuttajat sekä niiden analysointi
- otoskoon määrittely
- satunnaistamisessa käytettävä menetelmä.

33-kohtainen SPIRIT-tarkistuslista keskittyy siis suunnitelmien sisältöön pikemminkin kuin muotoiluun. Tarkistuslista suosittelee kattavaa kuvausta suunnitelluista toimenpiteistä, mutta ei määrää, miten tutkimus tulisi suunnitella tai toteuttaa. Tarjoamalla ohjeistusta keskeisistä sisällöistä SPIRIT-lausuma pyrkii helpottamaan korkealaatuisten tutkimussuunnitelmien laatimista. Suunnittelussa tarvitaan ylipäänsä tietoa tutkimuslainsäädännöstä, lupaprosesseista, eettisistä asioista, tilastotieteestä ja tutkimusaineiston keräämisestä. Useimmat tiedelehdet ovat ottaneet käyttöön SPIRIT-lausuman ja siihen liittyvän tarkistuslistan joko sellaiseenaan tai lehden omiin tarpeisiin parhaiten soveltuvin osin. Näin esimerkiksi läpinäkyvästi ja kattavasti raportoitu tutkimussuunnitelma, joka määrittelee tutkimusmenetelmät etukäteen, auttaa tunnistamaan jälkikäteen tehdyt muutokset, kuten päävastemuuttajan vaihtamisen. Ilmiö on valitettavan yleinen ja voi aiheuttaa harhaa interventioiden vaikutuksia koskevaan kirjallisuuteen.

Tilastollinen analyysisuunnitelma (*Statistical Analysis Plan, SAP*) on osa tutkimussuunnitelmaa tai siihen kiinteästi liittyvä erillinen dokumentti. Dokumenttia voi tutkimuksen biostatistikko tai tilastollisesta analyysistä vastaava henkilö täydentää aina satunnaistamiskoodin avaamiseen asti. Esimerkiksi psykososiaalisia interventioita tutkivissa vertailukokeissa sekä tutkimushenkilö että interventiota toteuttava henkilö tietävät, mihin satunnaistettuun ryhmään kyseinen tutkimushenkilö kuuluu, joten tutkimusta analysoivan henkilön on hyvä pysyä sokkona tutkimuksen laadun ylläpitämiseksi ainakin toissijaisten vastemuuttajien analysointiin asti (Hinkka-Yli-Salomäki, 2022). Monissa lehdissä satunnaistetun vertailukokeen tuloksiin pohjautuvaa käsikirjoitusta ei julkaista, mikäli tilastollista analyysisuunnitelmaa ei ole lähetetty käsikirjoituksen arviointiin lähettämisen yhteydessä. Tämän vuoksi tilastollisen analyysisuunnitelman tekemiseen on kehitetty erilaisia valmiita pohjia ja yksityiskohtaisia ohjeita helpottamaan tutkimusryhmän työtä (Yuan ym., 2019; Stevens ym., 2023). SAP sisältää yksityiskohtaista tietoa seuraavista:

- **Analysoitavan tutkimusaineiston määrittely**
 - Otetaanko hoitoaieanalyysiin ITT-periaatteen (*Intent-to-treat*) mukaan kaikki satunnaistetut tutkimushenkilöt vai poistetaanko siitä mahdollisesti joitakin henkilöitä etukäteen määritelyjen syiden vuoksi?
- **Puuttuvien havaintojen käsittely**
 - Käytetäänkö analyyseissa jotakin imputointimenetelmää (kuten vakioarvoilla korvaaminen, regressiomallinnus tai koneoppiminen) tai estimointimenetelmää, joka ei poista tutkimusaineiston analyysistä tutkimushenkilöitä, joilla on puuttuvia havaintoja (kuten suurimman uskottavuuden estimointiin perustuvat menetelmät)?
- **Vastemuuttajien muunnokset**
 - Käytetäänkö esimerkiksi logaritimuunnosta, jotta vaste noudattaisi paremmin normaali-jakaumaa, joka mahdollistaisi paremmin tarkemman ja voimakkaamman tilastollisen testin käyttämisen?
- **Vastemuuttajakohtaiset tilastolliset mallit**
- **Käytettävä merkitsevyytaso**
- **Osajoukkoanalyysit**
- **Yhdysvaikutusten ja kovariaattien määrittely.**

Tilastollisia analyysimenetelmiä on useita erilaisia

Satunnaistetussa vertailukokeessa vastemuuttajien lähtötasomittaukset tehdään aina ennen satunnaistamista. Näitä vastemuuttajien lähtötasomittauksia verrataan toteutetun intervention jälkeisiin mittauksiin, joita voi olla – koasetelman valinnan mukaan – yksi tai useampia. Kun analysoidaan intervention jälkeisten vastemuuttajien mittausten muutoksia lähtötasosta vertailuryhmissä ja niiden välillä, voidaan arvioida esimerkiksi mielenkiinnon kohteena olevan intervention vaikuttavuutta.

”Satunnaistetussa vertailukokeessa vastemuuttajien lähtötasomittaukset tehdään aina ennen satunnaistamista.”

Edelleen käydään satunnaistetun vertailukokeessa kerättävän tutkimusaineiston analysointiin liittyvää vilkasta keskustelua siitä, tulisiko vastemuuttajan (päävastemuuttaja ja toissijaiset vastemuuttajat) tilastolliseen malliin sisällyttää kyseisen vastemuuttajan lähtötason arvo kovariaattina (*baseline adjustment*) (Twisk ym., 2018). Lähtötason sisällyttämistä tilastolliseen malliin kovariaattina vastustavat tutkijat ovat sitä mieltä, että kaikki vertailuryhmien perustason erot johtuvat sattumasta, satunnaistamisen ”epäonnistumisesta”, eikä korjausta satunnaistamisen jälkeen tehtäviin ryhmien välisiin vertailuihin tarvita, tai se on jopa väärin. Tutkijat, jotka puoltavat kovariaatin käyttöä sen sijaan väittävät, että regressio kohti keskiarvoa-ilmion (*regression to the mean*) vuoksi lähtötason arvolla korjaaminen on välttämätöntä. Toisin sanoen: kun lähtötilanteessa olevat erot koe- ja vertailu- tai kontrolliryhmän välillä johtuvat satunnaisvaihteluista ja mittausvirheistä, on taipumus, että keskimääräinen arvo laskee ryhmässä, jonka alkukeskiarvo oli korkein, ja nousee ryhmässä, jonka alkukeskiarvo oli matalin. Tätä ilmiötä kutsutaan regressioksi kohti keskiarvoa. Vaikka vaikuttaa siltä, että yleinen mielipide on kääntynyt vasteen lähtötason sisällyttämiseen kovariaattina tilastollisessa mallissa, kirjallisuudessa on edelleen

monia satunnaistettuja vertailukokeita, joissa [vasteiden tilastolliset analyysit on tehty ilman lähtötason korjausta](#). Toisaalta esimerkiksi Stuart G. Pocockin mukaan lähtötason mukaan ottaminen kovariaatiksi tai sen poisjättäminen vaikuttaa käytännössä hyvin vähän tutkimuksen vastemuuttujien estimaatteihin ja käsittelyryhmien vertailuun, koska satunnaistaminen varmistaa yleensä hyvän tasapainon käsittelyryhmien välillä (Pocock, 2015). Tilannetta auttaisi, jos esimerkiksi raportoinnissa käytetty CONSORT-lausuma ohjeistaisi tutkijoita lähtötason korjausten tekemiseen (Hinkka-Yli-Salomäki, 2023).

Tavoitteet

Kovariaatin käytöstä tai ylipäättään lähtötason huomioimisesta tilastollisissa analyysissä on paljon väärinkäsityksiä. Tämän artikkelin tavoitteena on:

- 1) esittää erilaisia edistyneempiä tilastollisia analyysimenetelmiä interventioiden vaikutusten arvioimiseksi
- 2) havainnollistaa näitä menetelmiä esimerkkiaineiston kautta
- 3) antaa suosituksia tutkimusaineiston analysointiin.

Tarkastelu rajataan jatkuviin vastemuuttujiin.

AINEISTO JA MENETELMÄT

Esimerkkiaineisto

Esimerkkiaineistona käytetään satunnaistetun vertailukokeen päävastemuuttujan uhmakkuus- ja käytösongelmien määrää mittaavan kyselyn pistemäärää (CBCL (*Child Behavior Checklist*) eksternali-

soiva 1,5–5-vuotiaille). Esimerkkiaineiston tutkimuksessa alle kouluikäisten lasten huoltajat satunnaistettiin joko hoito-ohjelmaan tai kontrolliryhmään. Hoito-ohjelmassa huoltajille opetettiin vanhemmuuden taitoja, ja kontrolliryhmäläisten huoltajat saivat tietoa alle kouluikäisten lasten uhmakkuus- ja käytösongelmista. Taulukko 1 esittää uhmakkuus- ja käytösongelmia mittaavan kyselyn kuvaavia tunnuslukuja aikapisteittäin hoito- ja kontrolliryhmässä.

Analyysimenetelmät

Tässä artikkelissa tarkastellaan hoitovaikutusten arviointiin käytettyä kolmea analyysimenetelmää, joita sovellettiin esimerkkiaineistoon:

- 1) Toistomittausten kovarianssianalyysi (*Repeated measures analysis of covariance, ANCOVA*)
- 2) Toistomittausten analyysi (*Repeated measures analysis*)
- 3) Muutosten analyysi (*Analysis of changes*).

Toistomittauksilla tarkoitetaan samalta tutkittavalta kerättyjä toistuvia vastemuuttujien arvoja. Eri menetelmien tarkasteluissa oletuksena on, että vastemuuttujan mittausajankohtia on kolme: lähtötaso ennen satunnaistamista ja kaksi satunnaistamisen jälkeen. Edellä mainituista menetelmistä käytetään yleisesti myös nimitystä hierarkkiset lineaariset sekamallit (*hierarchical linear mixed models*), jossa hierarkkisuus viittaa kunkin käsittelyryhmän tutkittavien toistomittauksiin. Tutkittavat ovat siis ”pesiytyneet” käsittelyryhmien sisälle. Menetelmiä voidaan helposti laajentaa sisältämään useampia mittausajankohtia satunnaistamisen jälkeen. Toisinaan kirjallisuudessa samoista menetelmistä käytetään myös pelkästään nimitystä lineaariset sekamallit (*linear mixed models*).

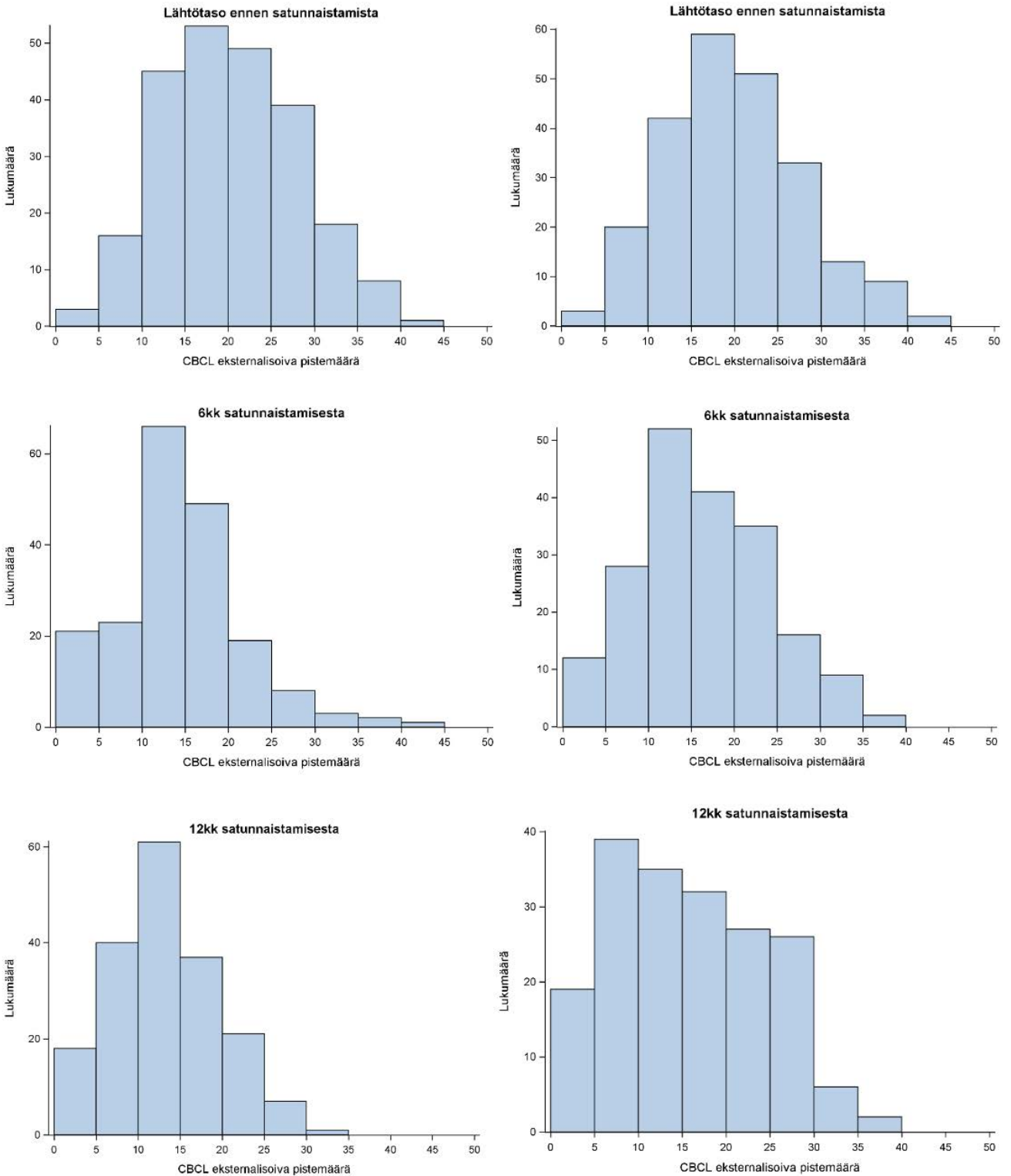
TAULUKKO 1.

Vastemuuttujan kuvailevat tunnusluvut hoito- ja kontrolliryhmissä aikapisteittäin.

Tilastollinen tunnusluku	Uhmakkuus- ja käytösongelmien määrää kuvaava pistemäärä (CBCL eksternalisoiva)	
	Hoitoryhmä	Kontrolliryhmä
Lähtötaso (ennen satunnaistamista)		
Keskiarvo	19,92	19,4
Keskivirhe	0,51	0,52
Kvartiiliväli	14–26	14–24
N	232	232
6 kuukautta satunnaistamisesta		
Keskiarvo	14,03	16,08
Keskivirhe	0,52	0,55
Kvartiiliväli	10–17	11–22
N	193	195
12 kuukautta satunnaistamisesta		
Keskiarvo	12,78	15,28
Keskivirhe	0,46	0,62
Kvartiiliväli	9–17	8–22
N	186	186

KUVIO 1.

Vastemuuttujan jakaumat kolmessa eri aikapisteessä: lähtötasossa sekä kuuden ja 12 kuukauden seurantamittauspisteessä hoito- ja kontrolliryhmissä.



Rajoitetun suurimman uskottavuuden estimointimenetelmän (*Restricted Maximum Likelihood*, REML) avulla saadaan mahdollisimman harhattomat estimaatit tilastollisen mallin parametreille, jotka liittyvät käsittelyryhmään, aikapisteisiin ja näiden yhdysvaikutuksiin sekä kovariaattiin. Tarkasteltaviin tilastollisiin malleihin voidaan sisällyttää sekä kiinteitä että satunnaisia tekijöitä. Esimerkkiaineistossa kiinteitä eli ei-satunnaisia tekijöitä (*fixed effect*), joiden havaittuihin arvoihin analyysin johtopäätökset voidaan yleistää, ovat käsittelyryhmä (hoito- ja kontrolliryhmä) ja aikapiste (lähtötaso, 6 kuukautta ja 12 kuukautta satunnaistamisesta). Satunnaisena tekijänä (*random effect*), jonka koko jakaumaan analyysin johtopäätökset voidaan yleistää, voidaan pitää tutkittavan osallistujan vaikutusta. Tyypillisesti kategoriset tekijät käsitellään kiinteinä tekijöinä. Kiinteitä ja satunnaisia tekijöitä sisältävää mallia kutsutaan tilastotieteen termein sekamalliksi (*mixed model*).

”Toistomittauksilla tarkoitetaan samalta tutkittavalta kerättyjä toistuvia vastemuuttujien arvoja.”

Kovarianssirakenteet

Kun mittauksia on useita samalta tutkimushenkilöltä, ei voida olettaa, että mittaukset olisivat toisistaan riippumattomia. Näin ollen peräkkäisten esimerkiksi *t*-testien suorittaminen ei ole sopivaa monitestaustilanteen takia ja on käytettävä toistetuille mittauksille soveltuvia menetelmiä (Hinkka-Yli-Salomäki, 2023). Riippuvuus huomioidaan matemaattisen varianssi- ja kovarianssirakenteen avulla. Kun riippuvuutta kuvaamaan sovitetaan kovarianssirakenne, jossa on vähemmän estimoitavia parametreja, kiinteiden tekijöiden estimaatteja voidaan tarkentaa (Littell ym., 2000). Jos aineistoon sovitettava kovarianssirakenne yksinkertaistaa liikaa henkilöiden sisäisiä ja välisiä vaihteluita, riski virheelliselle tilastolliselle päätökselle intervention vaikuttavuudesta kasvaa. Tällöin ei huomioida riittävästi aineiston ominaispiirteitä ja menetetään tilastollisesti merkitsevää informaatiota. Analysoitaessa aineistoa edistyneemmillä tilastollisilla ohjelmistoilla voidaan sovittaa kymmeniä erilaisia kovarianssirakenteita. Näistä yleisimpiä ovat

- heterogeeninen vakiovariانسsi -rakenne (*heterogeneous compound symmetry*, CSH)
- Toeplitz-rakenne
- ensimmäisen asteen autoregressiivinen rakenne (AR(1))
- vakiovariانسsi-rakenne (*compound symmetry*, CS).

Heterogeeninen vakiovariانسsi -rakenne tarkoittaa sitä, että jokaiselle aikapisteelle estimoidaan oma varianssinsa korrelaation pysyessä vakiona. Vaihtuva variانسsi kuulostaa järkevältä ajatukselta, koska voidaan olettaa, että ajallisesti lähellä toisiaan olevien mittausten varianssin vaihtelu olisi pienempää kuin kauempana toisistaan olevien. Toeplitz-rakenteessa oletetaan, että kaikkien aikapisteiden mittausten variانسsi on sama, mutta korrelaatio vaihtelee aikapisteen mukaan. Toisin sanoen kauempana olevien aikapisteen mittaukset ovat vähemmän korreloivia kuin lähellä olevien. AR(1)-rakenteessa korrelaatio on sama kahden vierekkäisen aikapisteen mittaauksissa ja vähenee eksponentiaalisesti siirryttäessä kauempana oleviin mittauksiin. Vakiovariانسsi-rakenteessa variانسien oletetaan olevan kaikkien aikapisteiden mittausten välillä yhtä suuria ja korrelaation sama aikapisteiden läheisyydestä riippumatta.

”Vakiovariانسsi-rakenteessa variانسien oletetaan olevan kaikkien aikapisteiden mittausten välillä yhtä suuria ja korrelaation sama aikapisteiden läheisyydestä riippumatta.”

Vakiovariانسsi-rakenne oli pitkään ainoa vaihtoehto toistettujen mittausten analysoimisessa. Yleisesti ottaen kiinteiden vaikutusten estimaatit, kuten hoito- ja kontrolliryhmän keskiarvojen välinen ero, voi olla sama tai lähes sama kovarianssirakenteesta riippumatta, mutta estimaattien keskivirheissä voi olla huomattavaa eroa. Siksi kovarianssirakenteeseen on hyvä kiinnittää huomiota.

TULOKSET

Kovarianssirakenteiden vertailu

Taulukossa 2 on vertailtu analyysimenetelmän 1 osalta yleisimpiä kovarianssirakenteita toisiinsa kahdella tavalla. Nämä tavat ovat uskottavuusosamäärätesti ja Akaiken informaatiokriteeri (*Akaike's information criterion*, AIC). Uskottavuusosamäärätestissä χ^2 -jakauman vapausasteet lasketaan vähentämällä monimutkaisemman eli enemmän estimoitavia parametreja sisältävän mallin parametrit yksinkertaisemman mallin parametrien lukumäärästä. Testisuureen arvo sitä vastoin saadaan vähentämällä monimutkaisemman mallin ($-2 \times \text{REML}$)-arvosta yksinkertaisemman mallin ($-2 \times \text{REML}$)-arvo. Testin merkitsevyys päätetään vertaamalla saatua testisuureen arvoa χ^2 -jakauman kriittiseen arvoon yllä esitetyllä vapausasteella. Toinen tapa parhaimman kovarianssirakenteen valitsemiseksi on Akaiken informaatiokriteeri. Paras kovarianssirakenne on se, jonka AIC-arvo on pienin. Vastaavia kriteereihin perustuvia tapoja on muitakin, mutta AIC-tapa on yleisin ja laskettavissa useimmilla tilasto-ohjelmilla.

Taulukossa 2 ilmoitetun uskottavuusosamäärätestin mukaan kovarianssimatriisin estimoitavien parametrien määrää ei saada pienennettyä, eli sovitetuista neljästä rakenteesta mitään ei voida valita vaarantamatta tilastollisen päättelyn luotettavuutta intervention vaikuttavuuden suhteen. Toisin sanoen: kun aineistoon sovitettua rakenteetonta kovarianssirakennetta verrataan kaikkiin muihin yksinkertaisimpiin rakenteisiin, saadaan testin tulokseksi merkitsevä *p*-arvo. AIC-arvoihin perustuva päättely tukee tätä, sillä pienin AIC-arvo (7942,7) saadaan, kun aineistoon ei soviteta mitään kovarianssirakennetta.

Hoitovaikutusten arviointi

Esimerkkiaineistossa hoitovaikutuksen arvioinnissa mielenkiinnon kohteena ovat seuraavat kolme vertailua käytösongelmien määrää kuvaavassa pistemäärässä hoito- ja kontrolliryhmien välillä (CBCL eksternalisoiva):

- 1) ero kuuden kuukauden kohdalla verrattuna lähtötasomittaukseen
- 2) ero 12 kuukauden kohdalla verrattuna lähtötasomittaukseen
- 3) ero 12 kuukauden kohdalla verrattuna kuuden kuukauden mittaukseen, jonka avulla tutkitaan hoidon mahdollista vaikutuksen pysyvyyttä pitkällä aikavälillä.

Kovarianssirakenteiden vertailu esimerkkiaineistossa.

Malli	Kovarianssirakenne	Tapa 1		χ^2 -arvo	Vapausasteet	Tapa 2	
		(-2×REML)-arvo	Parametrien lukumäärä			p -arvo ^a	AIC
1	Ei rakennetta (Unstructured)	7930,7	6				7942,7
2	Heterogeeninen vakiovarianssi	7947,9	4	17,2 (=7947,9-7930,7)	2 (= 6 - 4)	< 0.001	7955,9
3	Toeplitz	7943,3	3	13,4 (=7943,3-7930,7)	3 (= 6 - 3)	< 0.01	7949,3
4	Ensimmäisen asteen autoregressiivinen (AR(1))	7969,7	2	39,2 (=7969,7-7930,7)	4 (= 6 - 2)	< 0.001	7973,7
5	Vakiovarianssi	7951,5	2	20,8 (=7951,5-7930,7)	4 (= 6 - 2)	< 0.001	7955,5

^a p -arvon merkitsevyys saadaan vertaamalla χ^2 -jakauman kriittisiin arvoihin.

Toistomittausten kovarianssianalyysi (analyysimenetelmä 1)

Toistomittausten kovarianssianalyysissa aineistoon sovitettiin tilastollinen malli, jossa kiinteinä tekijöinä olivat

- käsittelyryhmä (hoito- ja kontrolliryhmä)
- aika (6 kuukautta ja 12 kuukautta satunnaistamisesta)
- käsittelyryhmän ja ajan yhdysvaikutus
- kovariaattina vastemuuttujan lähtötasomittaus ennen satunnaistamista.

Satunnaisena tekijänä pidettiin tutkittavan osallistujan vaikutusta. Hoitovaikutusten arviointiin käytettiin tilastollisen mallin käsittelyryhmän ja ajan yhdysvaikutuksen kontrasteja. Kontrastilla tarkoitetaan mekanismia, jolla saadaan halutut hypoteesien mukaiset vertailut testattua. Taulukossa 3 esitetään mallin mukaiset keskiarvot keskivirheineen jokaisessa aikapisteessä erikseen. Kuuden kuukauden kohdalla hoitoryhmässä ovat CBCL eksternalisoivan pisteet vähentyneet keskimäärin -2,5 pistettä (95 %:n luottamusväli -3,6–(-1,3)) verrattuna kontrolliryhmään (Taulukko 4). Tämä voidaan todentaa laskemalla taulukon 3 mallin mukaisista keskiarvoista: 13,7–16,2 = -2,5 pistettä. Vastaavasti 12 kuukauden kohdalla ero on hieman suurempi eli -2,9 pistettä (95 %:n luottamusväli (-4,1)–(-1,7)). Tulos saadaan todennettua myös taulukon 3 keskiarvoista: 12,6–5,5 = -2,9 pistettä.

”Kontrastilla tarkoitetaan mekanismia, jolla saadaan halutut hypoteesien mukaiset vertailut testattua.”

Lähtötason jälkeisissä aikapisteissä on jonkin verran puuttuvia havaintoja: 6 kuukauden kohdalla satunnaistamisen jälkeen hoitoryhmässä on 39 (16,8 %) ja kontrolliryhmässä 37 (15,9 %) puuttuvaa. Käsittelyryhmien välinen puuttuvien havaintojen ero ei ole tilastollisesti merkitsevä (Fisherin tarkka testi (Fisher's exact test),

p -arvo = 0,90). Vastaavasti 12 kuukauden kohdalla vastaavat luvut ovat molemmissa ryhmissä 46 (19,8 %). Puuttuvien määrän ollessa suhteellisen pieni voidaan käyttää REML-estimointimenetelmää, jolla saadaan kaikki tutkimushenkilöt mukaan riippumatta siitä, onko heillä puuttuvia arvoja vai ei (Hinkka-Yli-Salomäki, 2023).

Toistomittausten analyysi (analyysimenetelmä 2)

Toistomittausten analyysissa aineistoon sovitettiin tilastollinen malli, jossa kiinteinä tekijöinä olivat

- käsittelyryhmä (hoito- ja kontrolliryhmä)
- aika (lähtötaso ennen satunnaistamista, 6 kuukautta ja 12 kuukautta satunnaistamisesta)
- käsittelyryhmän ja ajan yhdysvaikutus.

Satunnaisena tekijänä pidettiin tutkittavan osallistujan vaikutusta. Hoitovaikutusten arviointiin käytettiin tilastollisen mallin käsittelyryhmän ja ajan yhdysvaikutuskontrasteja. Taulukossa 3 esitetään mallin mukaiset keskiarvot keskivirheineen jokaisessa aikapisteessä erikseen. Kuuden kuukauden kohdalla hoitoryhmässä ovat CBCL eksternalisoivan pisteet vähentyneet keskimäärin 2,7 pistettä (95 %:n luottamusväli 1,4–4,0) verrattuna kontrolliryhmään (Taulukko 4). Tämä voidaan todentaa laskemalla taulukon 3 mallin mukaisista keskiarvoista: (19,9–13,9) - (19,4–16,1) = 2,7 pistettä. Vastaavasti 12 kuukauden kohdalla ero on hieman suurempi eli 3,1 pistettä (95 %:n luottamusväli 1,8–4,5). Tulos saadaan lasketta taulukon 3 keskiarvoista: (19,9–12,8) - (19,4–15,4) = 3,1 pistettä.

Muutosten analyysi (analyysimenetelmä 3)

Muutosten analyysissa aineistoon sovitettiin tilastollinen malli, jossa kiinteinä tekijöinä olivat

- käsittelyryhmä (hoito- ja kontrolliryhmä)
- muutos (lähtötasosta kuuteen kuukauteen ja lähtötasosta 12 kuukauteen satunnaistamisesta)
- käsittelyryhmän ja ajan yhdysvaikutus
- kovariaattina lähtötasomittauksen arvot.

Uhmakkuus- ja käytösongelmien määrää kuvaavan pistemäärän (CBCL eksternalisoiva) keskiarvot ja keskivirheet aikapisteissä esitetynä kolmella eri analyysimenetelmällä analysoituna.

Analyysimenetelmä	Uhmakkuus- ja käytösongelmien määrää kuvaava pistemäärä (CBCL eksternalisoiva)					
	Hoitoryhmä			Kontrolliryhmä		
	Lähtötaso Keskiarvo (SE) ^c	6 kk ^a Keskiarvo (SE) ^c	12 kk ^b Keskiarvo (SE) ^c	Lähtötaso Keskiarvo (SE) ^c	6 kk ^a Keskiarvo (SE) ^c	12 kk ^b Keskiarvo (SE) ^c
Toistomittausten kovarianssianalyysi	-	13,7 (0,4)	12,6 (0,4)	-	16,2 (0,4)	15,5 (0,4)
Toistomittausten analyysi	19,9 (0,5)	13,9 (0,5)	12,8 (0,5)	19,4 (0,5)	16,1 (0,5)	15,4 (0,5)
Muutosten analyysi	-	5,8 (0,4)	7,0 (0,4)	-	3,3 (0,4)	4,0 (0,4)

^a 6 kuukautta satunnaistamisesta

^b 12 kuukautta satunnaistamisesta

^c Mallin mukainen keskiarvo ja keskivirhe

Satunnaisena tekijänä pidettiin tutkittavan osallistujan vaikutusta. Hoitovaikutusten arviointiin käytettiin tilastollisen mallin käsittelyryhmän ja ajan yhdysvaikutuskontrasteja. Taulukossa 3 esitetään mallin mukaiset keskiarvot keskivirheineen kuuden ja 12 kuukauden aikapisteissä erikseen. Kuuden kuukauden kohdalla hoito- ja kontrolliryhmän ero CBCL eksternalisoivan keskimääräisissä pisteissä on 2,5 pistettä (95 %:n luottamusväli 1,3–3,6) verrattuna kontrolliryhmään (taulukko 4). Tämä voidaan todentaa laskemalla taulukon 3 mallin mukaisista keskiarvoista: 5,8–3,3 = 2,5 pistettä. Vastaavasti 12 kuukauden kohdalla ero on hieman suurempi eli 3,0 pistettä (95 %:n luottamusväli 1,7–4,1). Tulos saadaan todennettua myös taulukon 3 keskiarvoista: 7,0–4,0 = 3,0 pistettä.

POHDINTA

Aiemmin on ajateltu, että kovariaattien eron todentamisessa hoito- ja kontrolliryhmien välillä tilastollisella testaamisella ei ole merkitystä, jos satunnaistaminen ja mahdollinen sokkoutus on tehty riittävän laadukkaasti. Nykyään ajatellaan, että jos jatkuvan vastemuuttujan lähtötaso on tiedossa, se pitäisi sisällyttää malliin kovariaattina (Twisk ym., 2018; European Medicines Agency, 2015). On siis perusteltua suositella toistomittausten kovarianssianalyysin käyttöä toistomittauksia sisältävässä satunnaistetussa vertailukokeessa. Se tarjoaa pienen lisän tilastolliseen voimaan ilman lisäkustannuksia ja vain vähäisellä tilastollisella vaivalla – joten miksi jättää tällainen mahdollisuus käyttämättä?

”Muutosten analyysi antaa saman tuloksen kuin toistomittausten kovarianssianalyysi, mikäli aineistossa ei ole puuttuvia havaintoja.”

Menetelmän lähtötasomittaus, joka on tehty ennen intervention aloittamista, käsitellään tilastollisessa mallissa kovariaattina. Muutos lähtötasosta on riippuvainen itse lähtötasosta. Toisin sanoen henkilöillä, joilla on korkea lähtötaso mitattavassa vastemuuttu-

jassa, havaitaan useimmiten suurempaa pudotusta kuin tutkittavilla osallistujilla, joilla on alhainen lähtötaso jo valmiiksi. Ilmiötä kutsutaan palautumiseksi keskiarvoa kohti (*regression to the mean*). Satunnaistettujen kontrolloitujen tutkimusten (RCT) suunnittelussa ja toistuvasti mitattujen jatkuvien vastemuuttujien analysoinnissa on suositeltavaa käyttää tilastollisesti edistyneitä menetelmiä. Muutosten analyysi antaa saman tuloksen kuin toistomittausten kovarianssianalyysi, mikäli aineistossa ei ole puuttuvia havaintoja. Rajoitetun suurimman uskottavuuden estimointimenetelmä, joka löytyy nykyään yleisimmistä tilasto-ohjelmista, pystyy luotettavasti estimoimaan, vaikka aineistossa olisi puuttuvia havaintoja. Nyrkkisääntönä pidetään yleisesti niin sanottua 30 prosentin sääntöä eli satunnaisesti puuttuvia havaintoja sallitaan aineistossa enintään 30 prosenttia. Puuttuvat havainnot voidaan luokitella seuraavasti:

- 1. Täysin satunnaiset (MCAR, missing completely at random):** puuttuvan arvon todennäköisyys ei riipu havaituista eikä havaitsemattomista mittauksista.
- 2. Satunnaiset (MAR, missing at random):** puuttuvan arvon todennäköisyys riippuu havaituista, mutta ei havaitsemattomista mittauksista.
- 3. Informatiiviset (Informative missingness):** puuttuvan arvon todennäköisyys riippuu havaitsemattomista mittauksista.

Edellä esitettyjen kolmen analyysimenetelmän lisäksi voidaan käyttää erilaisia regressioanalyysiin pohjautuvia menetelmiä. Näissä aineistoon sovitetaan lineaarisia ja epälineaarisia regressiokäyriä ajan funktiona, ja vertailu hoito- ja kontrolliryhmien vaikuttavuuksista tehdään tietyissä aikapisteissä, kuten esimerkiksi esimerkkiaineistossa kuuden ja 12 kuukauden kohdalla satunnaistamisesta. Tavallisesti satunnaistetuissa vertailukokeissa mielenkiinto on ryhmien välisessä vertailussa eikä niinkään ilmiöiden mallintamisessa, joten toistomittausten kovarianssianalyysi, toistomittausten analyysi ja muutosten analyysi ovat suositeltavimpia menetelmiä.

Toistettujen mittausten analysoinnissa ei ole sopivaa analysoida aineistoa esimerkiksi yksisuuntaisella varianssianalyysillä, koska malli olettaa eri aikapisteissä tehdyt mittaukset riippumattomiksi.

Toinen epäsopeva tapa on vertailla käsittelyryhmien eroja kussakin aikapisteessä erikseen esimerkiksi *t*-testein. Aiemmin kun teoria ja siihen liittyvät tilastolliset ohjelmit eivät olleet vielä kovin kehittyneitä, aineistoa yksinkertaistettiin, jotta riippuvuudesta päästiin eroon. Yksinkertaistaminen sisälsi esimerkiksi keskiarvon laskemisen yli aikapisteiden, eli ennen satunnaistamista tehdyistä mittauksista laskettiin keskiarvo, jota verrattiin satunnaistamisen jälkeen mitattujen keskiarvoon. Toinen lähes yhtä yleinen tapa oli käyrän alapuolelle jäävän pinta-alan (AUC, *area under curve*) laskeminen, jolloin tutkittavien toistetut mittaukset supistuivat yhteen arvoon, jolloin voitiin verrata käsittelyryhmien AUC-keskiarvoja toisiinsa. Erilaisten yksinkertaistamistapojen eli yhteenvetolukujen käytön suurimpana ongelmana on, että ajan vaikutuksen ominaispiirteet jäävät näissä huomioimatta.

”Tavallisesti satunnaistetuissa vertailukokeissa mielenkiinto on ryhmien välisessä vertailussa eikä niinkään ilmiöiden mallintamisessa, joten toistomittausten kovarianssianalyysi, toistomittausten analyysi ja muutosten analyysi ovat suositeltavimpia menetelmiä.”

Kun verrataan toistomittausten analyysia toistomittausten varianssianalyysiin (*repeated measures* ANOVA, RM ANOVA) tai kovarianssianalyysiin (*repeated measures* ANCOVA, RM ANCOVA), niin kaikilla tavoilla päädytään samaan tai lähes samaan tulokseen yksinkertaisessa ennen–jälkeen-koeasetelmassa.

Oletuksena tässä on, että

1. aineistossa ei ole puuttuvia havaintoja
2. mallin vaikutukset ovat kaikki kiinteitä tekijöitä
3. molempien mallien jäännöstermi noudattaa normaalijakaumaa.

Todellisuudessa hyvin harvoin ihmisillä tehtävissä tutkimuksissa tutkijat pystyvät keräämään täydellisen aineiston vailla puuttuvia havaintoja. Eläinkokeissa sen sijaan tämä on useimmiten mahdollista.

Tässä artikkelissa esitetyt kolme analyysitapaa ovat ylivertaisia verrattuna erilaisiin RM AN(C)OVA -malleihin seuraavissa tapauksissa:

1. Aineistossa on puuttuvia havaintoja. RM AN(C)OVA -malleissa henkilöt, joilla on vähintään yksi puuttuva havainto, putoavat pois analyysistä.
2. Koeasetelmassa on useita seurantamittauksia, jolloin kovarianssirakenteen mallittamisella saadaan estimaattien tarkkuutta parannettua.
3. Mittauksia ei tehdä tasavälein. Satunnaistetuissa vertailukokeissa harvoin tehdään kaikki mittaukset esimerkiksi tarkalleen viikon välein, vaan intervention jälkeen mittausajankohdat ovat usein lähempänä toisiaan kuin pitemmän ajan seurantamittaukset.
4. Koeasetelma sisältää klustereita eli ryppäitä. Tällöin satunnaistaminen tehdään usein ryssäsatunnaistamisena (cluster randomisation), jossa esimerkiksi satunnaistetaan isompia yksiköitä, kuten koulut ja päiväkodit, yksittäisten tutkittavien sijaan.
5. Mallissa on mukana satunnaisia tekijöitä, kuten tutkittavat. Tällöin katsotaan, että tulokset ovat yleistettävissä koko perusjoukkoa koskevaksi.

Ren kollegoineen (2022) tarkasteli yli 200:a satunnaisesta vertailukokeesta tehtyä raporttia, ja vain alle puolessa käytettiin toistomittauksille soveltuvia tilastollisia menetelmiä. Lähtötasomittausta kovariaattina tilastollisissa malleissa käytti vain murto-osa.

TAULUKKO 4.

Uhmakuus- ja käytösongelmien määrää kuvaavan pistemäärien (CBCL eksternalisoiva) vertailu hoito- ja kontrolliryhmien välillä 6 ja 12 kuukautta satunnaistamisen jälkeen kolmella eri analyysimenetelmällä analysoituna.

Analyysimenetelmä	Uhmakuus- ja käytösongelmien määrää kuvaava pistemäärä (CBCL eksternalisoiva) Hoito- vs. kontrolliryhmä					
	Lähtötasomittaus vs. 6 kk ^a :n mittaus Keskiarvo ^c		Lähtötasomittaus vs. 12 kk ^b :n mittaus Keskiarvo ^c		6 kk ^a :n mittaus vs. 12 kk ^b :n mittaus Keskiarvo ^c	
	(95 % lv)	<i>p</i> -arvo	(95 % lv)	<i>p</i> -arvo	(95 % lv)	<i>p</i> -arvo
Toistomittausten kovarianssianalyysi	-2,5 (-3,6–(-1,3))	< 0.001	-2,9 (-4,1–(-1,7))	< 0.001	0,5 (-0,7–1,6)	0,43
Toistomittausten analyysi	2,7 (1,4–4,0)	< 0.001	3,1 (1,8–4,5)	< 0.001	0,4 (-0,7–1,6)	0,44
Muutosten analyysi	2,5 (1,3–3,6)	< 0.001	3,0 (1,7–4,1)	< 0.001	-0,5 (-1,6–0,7)	0,43

^a 6 kuukautta satunnaistamisesta

^b 12 kuukautta satunnaistamisesta

^c Käsittelyryhmien välinen mallin mukainen ero

lv: luottamusväli

Yli puolet tilastollisesti edistyneitä menetelmiä, kuten toistomittausten analyysia, käyttäneistä tutkimuksista ei määritellyt tarkasti tilastollista mallia koskevia yksityiskohtia, kuten esimerkiksi korrelaatorakenteen oletuksia, mukaan otettuja muuttujia tai lähtötasomittausten mukaan ottamista kovariaattina. Vaikka tilastollisesti edistyneitä menetelmiä on suositeltu toistomittausten analysointiin jo 1990-luvulta lähtien, on edelleen epäselvää, miten nykyiset tutkimukset käytännössä käsittelevät tällaisia aineistoja. Ren tunnisti useita keskeisiä metodologisia puutteita eri tieteellisten julkaisujen välillä, joten toistuvasti mitattujen jatkuvien vasteiden tilastollisen analyysin laatua olisi saatava nostettua laajentamalla edistyneempien menetelmien käyttöönottoa.

”Satunnaistettujen vertailukokeiden suunnittelussa ja toistuvasti mitattujen jatkuvien vastemuuttujien analysoinnissa on erittäin suositeltavaa käyttää tilastollisesti edistyneitä menetelmiä.”

Yhteenvetona erityisesti tutkijoille kohdistaisin, että satunnaistettujen vertailukokeiden suunnittelussa ja toistuvasti mitattujen jatkuvien vastemuuttujien analysoinnissa on erittäin suositeltavaa käyttää tilastollisesti edistyneitä menetelmiä. Toistomittausten analyysimenetelmien ja tilastollisten yksityiskohtien (esimerkiksi korrelaatorakenteen oletusten, malliin sisällytettävien muuttujien ja lähtötasomittauksen kovariaatin) etukäteismäärittely on perusteltua. Tämä parantaa tilastollisten analyysien metodologista laatua ja edistää tutkimuksen läpinäkyvyyttä. ●

Tätä tutkimusartikkelia ei ole vertaisarvioitu.

AVAINSANAT:

satunnaistettu vertailukoe, RCT, analysointi, toistomittaus, jatkuva vastemuuttuja

Saapunut: 24.4.2025

Hyväksytty: 30.5.2025

LÄHTEET:

- Chan, A.-W., Tetzlaff, J. M., Altman, D. G., Laupacis, A., Gøtzsche, P. C., Krleža-Jerić, K., ... & Moher, D. (2013). SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. *Ann Intern Med*, 158, 200–207. <https://doi.org/10.7326/0003-4819-158-3-201302050-00583>
- European Medicines Agency. (2015). Guideline on adjustment for baseline covariates in clinical trials. EMA/CHMP/295050/2013. Haettu osoitteesta <https://www.ema.europa.eu/en/adjustment-baseline-covariates-clinical-trials-scientific-guideline>
- Hinkka-Yli-Salomäki, S. (2021). Vaikuttavuustutkimus: Satunnaistettu vertailukoe tulee suunnitella huolella. *Kasvun tuki -aikakauslehti*, 1(2). <https://doi.org/10.61259/kt.128200>
- Hinkka-Yli-Salomäki, S. (2022). Vaikuttavuustutkimus: Satunnaistetun vertailukokeen laadukkaan raportoinnin edellytykset. *Kasvun tuki -aikakauslehti*, 2(1). <https://doi.org/10.61259/kt.128825> Hinkka-
- Hinkka-Yli-Salomäki, S. (2023). Vaikuttavuustutkimus: Mitä tulee huomioida analysoitaessa satunnaistetun vertailukokeen aineistoa?. *Kasvun tuki -aikakauslehti*, 3(1). <https://doi.org/10.61259/kt.130461>
- Littell, R. C., Pendergast, J. & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statist. Med*, 19, 1793–1819. [https://doi.org/10.1002/1097-0258\(20000715\)19:13<1793::AID-SIM482>3.0.CO;2-Q](https://doi.org/10.1002/1097-0258(20000715)19:13<1793::AID-SIM482>3.0.CO;2-Q)
- Pocock, S. J., Clayton, T. C., Gregg, W. & Stone, G. W. (2015). Design of Major Randomized Trials: Part 3 of a 4-Part Series on Statistics for Clinical Trials. *J Am Coll Cardiol*, 66, 2757–66. <http://dx.doi.org/10.1016/j.jacc.2015.10.036>
- Ren, Y., Yuanjin, Z., Yulong, J., Yunxiang, H., Minghong, Y., Ling, L., ... & Xin, S. (2022). Analyses of repeatedly measured continuous outcomes in randomized controlled trials needed substantial improvements. *Journal of Clinical Epidemiology*, 143, 105–117. <https://doi.org/10.1016/j.jclinepi.2021.12.007>
- Stevens, G., Dolley, S., Mogg, R. & Connor, J. T. (2023). A template for the authoring of statistical analysis plans. *Contemp. Clin. Trials*, 34, 101100. <https://doi.org/10.1016/j.conctc.2023.101100>
- Tunn, R., Boutron, I., Chan, A.-W., Collins, G. S., Hróbjartsson, A., Moher, D., ... & Hopewell, S. (2024). Methods used to develop the SPIRIT 2024 and CONSORT 2024 Statements. *JCE*, 169, 111309. <https://doi.org/10.1016/j.jclinepi.2024.111309>
- Twisk, J., Bosman, L., Hoekstra, T., Rijnhart, J., Welten, M. & Heymans, M. (2018). Different ways to estimate treatment effects in randomised _controlled trials. *Contemp. Clin. Trials*, 10, 80–85. <https://doi.org/10.1016/j.conctc.2018.03.008>
- Yuan, I., Topjian, A. A., Kurth, C. D., Kirschen, M. P., Ward, C. G., Zhang, B. & Mensinger, J. L. (2019). Guide to the statistical analysis plan. *Pediatric Anesthesia*, 29, 237–242. <https://doi.org/10.1111/pan.13576>