

Vaikuttavuustutkimus:

Mitä tulee huomioida analysoitaessa satunnaistetun vertailukokeen aineistoa?

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)

- Tutkimussuunnitelmassa tai viimeistään ennen kuin avataan satunnaistamiskoodi eli tieto siitä, millä menetelmällä kutakin tutkimushenkilöä hoidetaan, on määritettävä ennakkoon kovariaatit eli selittävät muuttujat, puuttuvan tiedon käsittelymenetelmät, herkkyysanalyysit, osajoukkoanalyysit ja tilastollisen hypoteesien järjestys.
- Perustason kovariaatti voidaan huomioida analyysivaiheessa sisällyttämällä se tilastolliseen malliin tai käyttämällä esimerkiksi ositettua satunnaistamista.
- Puuttuvan tiedon huomiotta jättäminen rikkoo hoitoaieperiaatetta, jonka mukaan analyysihin sisällytetään kaikki satunnaistetut tutkittavat.
- Monivertailuongelman välttämiseksi satunnaistetuissa vertailukokeissa kannattaa käyttää vain yhtä päävastetta. Mikäli tämä ei ole mahdollista, on käytettävä monivertailun huomioivia menetelmiä.
- Osajoukkoanalyysien tulokset ovat tärkeä esittää estimaattien ja luottamusvälien avulla, sillä pelkkä käsittelyryhmän ja kovariaatin yhdysvaikutuksen p-arvon esittäminen ei yksin riitä.



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

tojen käsittelyä, monivertailun huomioimista tilastollisissa analyysissä sekä osajoukoista tehtäviä johtopäätöksiä.

KOVARIAATTIEN KÄYTTÄMINEN TILASTOLLISISSA MALLEISSA

Tutkittavien satunnaistamisella pyritään estämään tutkittavien valikoitumista koe- ja vertailuryhmiin ja näin välttämään siitä aiheutuva valintaharhaa (Hinkka-Yli-Salomäki, 2022a). Satunnaistamisen tarkoituksena on, että tutkittavat jakaantuisivat ryhmiin täysin satumanvaraisesti. Tällöin tutkittavien perusominaisuuksien jakaumat, kuten sukupuoli-, ikä- ja koulutusjakaumat eivät välttämättä ole samanlaiset eri koe- ja vertailuryhmissä.

Perustason (*baseline*) kovariaatti määritellään muuttujaksi, joka mitataan yleensä ennen satunnaistamista ja ennen kuin tutkittava aloittaa hänelle kohdennetun (koe- tai vertailu)intervention ja jonka odotetaan vaikuttavan analysoitavaan päävasteeseen eli tärkeimpään vastemuuttujaan. Perustason kovariaatteja on erilaisia, ja niiden luonne riippuu tutkimuksen kontekstista. Ne voivat olla

- 1) demografisia muuttujia, kuten sukupuoli tai ikä
- 2) sairauden ominaisuuksia, kuten kesto tai vakavuus
- 3) mahdollisia muita prognostisia tekijöitä, joille on esimerkiksi jokin fysiologinen peruste
- 4) tutkimuksen suorittamiseen liittyviä tekijöitä, kuten tutkimuskeskus, jossa tutkittava osallistuu tutkimukseen tai interventiota antava tutkija.

Vaikuttavuustutkimusten kultaisena standardina pidetään satunnaistettuja vertailukokeita. Ne ovat yleisesti parhaisi hyväksytyjä koeasetelmia, kun verrataan kahden tai useamman intervention eli menetelmän vaikuttavuutta. Tämä pitää kuitenkin paikkansa vain, jos vaikuttavuustutkimus on asianmukaisesti suunniteltu, suoritettu ja raportoitu (Moher ym., 2010). Tämän artikkelin tavoite on lisätä ymmärrystä ja selvittää tilastollisia käsitteitä ja menettelyjä satunnaistettujen vertailukokeiden analyysien tavallisimmissa kysymyksissä. Artikkelin käsittelee perustason kovariaattien käyttöä, puuttuvien havain-

Yleinen kovariaatti on myös päävasteen perustason arvo, kuten esimerkiksi masennuskyselyyn tai yksinäisyyden kokemisen yksittäinen pistemäärä tutkimuksen alussa tai ennen satunnaistamista. (European Medicines Agency, 2015.)

Perustason kovariaatteja voidaan tarkastella satunnaistetuissa vertailukokeissa kahdessa vaiheessa: ne voidaan huomioida satunnaistamisprosessissa tyypillisesti käyttämällä ositettua satunnaistamista tai ne voidaan sisällyttää tekijöinä tilastollisiin malleihin, jolloin niiden vaikutus huomioidaan koe- ja vertailuryhmien välisissä vertailuissa (European Medicines Agency, 2015). Perustason kovariaatit voidaan sisällyttää tilastollisiin malleihin monella eri tavalla, jonka valinta riippuu kovariaattien, päävasteiden ja toissijaisten vasteiden luonteesta. Yleisesti käytettyjä menetelmiä ovat

- kvantitatiivisista vasteista kovarianssianalyysi (*analysis of covariance*, ANCOVA)
- kaksi- tai moniluokkaisista vasteista logistinen regressioanalyysi
- ajallisia vasteista (time-to-event) Coxin regressioanalyysi ja ositetut (stratifoidut) analyysit (esimerkiksi Cochran-Mantel-Haenszel-testi binääristen vasteiden tapauksessa).

Seuraavaksi tarkastellaan tilanteita, joissa perustason kovariaatit tulisi sisällyttää ensisijaisiin eli päävasteen analyysiin sekä esitetään, miten tulokset tulisi raportoida ja tulkita. Lähtökohtaisesti kovariaatit on aina määrittävä etukäteen tutkimussuunnitelmassa.

”Erytisen harmillista on, jos ero eli ryhmien välinen epätasapaino suosii koeryhmää eli sitä tutkittavien ryhmää, jossa intervention vaikutus halutaan todentaa.”

Satunnaistamisen odotetaan tasapainottavan koe- ja vertailuryhmät kovariaattien suhteen. Käytännössä ei ole kuitenkaan harvinaista, että tutkimusryhmien välillä havaitaan eroa kovariaattien arvoissa satunnaistamisen jälkeen. Erytisen harmillista on, jos ero eli ryhmien välinen epätasapaino suosii koeryhmää eli sitä tutkittavien ryhmää, jossa intervention vaikutus halutaan todentaa. Esimerkiksi tutkittaessa uutta opiskelijoille suunnattua stressinhallintamenetelmää havaitaan analysointivaiheessa, että koeryhmän keskimääräinen koettu stressi alkutilanteessa on huomattavasti suurempi kuin vertailuryhmän vastaava. Näin uudella menetelmällä voidaan mahdollisesti helpommin osoittaa koeryhmässä suurempi vaikutus, koska lähtötaso on koeryhmää korkeampi. Kun koettu stressi on jo valmiiksi matalalla tasolla, on sen laskeminen ehkä hankalampaa ja epävarmempaa kuin jos koettu stressitaso olisi alkutilanteessa hyvin korkea.

Tilastollisella testaamisella kovariaattien eron todentamisessa koe- ja vertailuryhmien välillä ei ole merkitystä, jos satunnaistaminen ja mahdollinen sokkoutus on tehty riittävän laadukkaasti (European Medicines Agency, 2015). Sokkoutamisessa tutkittavat eivät saa tietää, kuuluvatko he koe- vai vertailuryhmään. Kovariaatit esitetään mahdollisten epätasapainojenkin vallitessa koe- ja vertailuryhmissä ja havaitut eroavaisuudet voidaan käsitellä satunnaisina ilmiöinä. Kannattaa huomioida, että mikäli tutkittavien jakamista koe- ja vertailuryhmiin ei tehdä satunnaisesti, ei valiko-

tumisesta johtuvaa harhaa voida korjata jälkikäteen tilastollisella analyysillä (Hinkka-Yli-Salomäki, 2022b).

Ositettu satunnaistaminen

Ositusta (*stratification*) voidaan käyttää varmistamaan etukäteen tasapaino tutkimusryhmien välillä. Sitä voidaan käyttää hallinnollisista syistä esimerkiksi silloin, kun terapiaryhmiin osallistuvien perheiden tai rinnakkaisluokkien määrä koulussa on rajoitettu, eli niillä on tietty enimmäismäärä. Ositus saattaa parantaa intervention vaikutuksen arvioinnin tehokkuutta erityisesti pienissä tai kohtalaisen kokoisissa kokeissa. Monikeskustutkimukset ositetaan usein käytännön syistä tai kun keskusten oletetaan olevan sekoituneita tunnettuihin tai tuntemattomiin ennustaviin tekijöihin. Monikeskustutkimuksella tarkoitetaan tutkimusta, joka tehdään useammassa kuin yhdessä paikassa, ja paikat ovat esimerkiksi maantieteellisesti kaukana toisistaan (esimerkiksi Helsinki ja Kokkola) tai organisaationaalisesti erillään (esimerkiksi erikoissairaanhoidon poliklinikka ja lastenneuvola). Jos keskusten oletetaan olevan rekrytointimäärältään pieniä, niin ositus voidaan laajentaa esimerkiksi maantieteellisesti koskemaan useita keskuksia. Tästä esimerkkinä ovat yksittäisten terveyskeskusten yhdistäminen hyvinvointialueiden sisällä tai koulupiirien käyttäminen yksittäisten koulujen sijaan. Osituksessa käytetty muuttuja sisällytetään kovariaattina aina päävasteiden ja toissijaisten vasteiden tilastollisiin malleihin, paitsi jos ositus tehdään puhtaasti hallinnollisista syistä (European Medicines Agency, 2015).

Muuttujat, joiden tiedetään liittyvän vahvasti tai ainakin kohtalaisesti päävasteeseen, ja muuttujat, joilla on vahva perustelu tällaiselle yhteydelle, tulisi sisällyttää kovariaatteina ensisijaisiin analyysiin. Tällä perusteella valitut muuttujat tulee määrittää etukäteen tutkimussuunnitelmassa. Jälkikäteen havaittua perustason epätasapainoa koe- ja vertailuryhmien välillä, esimerkiksi jonkin taustamuuttujan suhteen, ei pidetä riittävänä syynä sisällyttää kyseistä muuttujaa kovariaattina ensisijaisiin analyysiin (Egbewale, 2015; European Medicines Agency, 2015). Sen sijaan herkkyyksianalyysien tekeminen ottamalla mukaan koe- ja vertailuryhmien välillä epätasapainossa oleva perustason muuttuja kovariaattina voi olla suureksi avuksi arvioitaessa ensisijaisten analyysien tulosten luotettavuutta.

Muuttujia, jotka on mitattu satunnaistamisen jälkeen, ei pitäisi sisällyttää kovariaatteina ensisijaisiin analyysiin, sillä interventio on voinut vaikuttaa muuttujan arvoihin (European Medicines Agency, 2015). Jos jatkuvan päävasteen perustason arvo (ns. lähtötaso) on saatavilla, se kannattaa ottaa mukaan kovariaattina. Tämä pätee riippumatta siitä, määritelläänkö päävaste muutokseksi lähtötasosta (*change from baseline*) vai tuloksena tietyssä aikapisteessä.

Esimerkkinä tästä on tutkimus, jossa tutkitaan uuden masennuksen hoitoon tarkoitetun kaksi viikkoa kestävä intervention vaikutusta. Vasteena on masennuskyselyyn pistemäärä, joka mitataan vaikutuksen arvioimiseksi ennen satunnaistamista ja 11 viikkoa satunnaistamisen jälkeen. Vertailuryhmä saa perusterveydenhuollon tarjoaman tavallisen hoidon. Muutos lähtötasosta -analyysi tarkoittaa tässä esimerkissä, että lasketaan lähtötalanteen ja 11 viikon mittauksen keskimääräinen erotus koe- ja vertailuryhmissä ja verrataan ryhmien erotusten arvoja toisiinsa. Tulos tietyssä aikapisteessä -analyysi vastaavasti tarkoittaa, että verrataan 11 viikon pistemääriä koe- ja vertailuryhmien välillä.

Kun vasteena on muutos lähtötasosta, niin analyysin tulokset, jossa perustason arvo on mukana kovariaattina, ovat yleensä

tarkempia kuin analyysin tulokset, jossa perustason arvo ei ole mukana. Tämä johtuu siitä, että hajonta on pienempää. Clifton on osoittanut, että molemmissa tapauksissa vasteen ollessa jatkuva kovarianssianalyysi on luotettava menetelmä. Kovarianssianalyysi korjaa mahdollisen epätasapainon jonkin perustason muuttujassa tutkimusryhmien välillä, ja sillä on suurempi tilastollinen voima kuin muilla menetelmillä (Clifton & Clifton, 2019; Vickers & Altman, 2001). Lähtötason käyttö kovariaattina lineaarisessa regressioanalyysissä antaa identtiset tulokset, oli vaste sitten muutos lähtötasosta tai varsinainen tulos. Näin ollen, jos sopiva korjaus tehdään, niin vasteen analyysin valinta muuttuu yksinomaan tulkintakysymykseksi.

”Kovariaattien käyttö on aina perusteltava.”

Kovariaattien käyttö on aina perusteltava, ja yleensä rajoitutaan käyttämään vain muutamaa kovariaattia ensisijaisissa analyyseissa. Aiemman tiedon puuttuessa jatkuvan kovariaatin ja vasteen väliselle suhteelle voidaan olettaa yksinkertainen funktionaalinen muoto (European Medicines Agency, 2015). Esimerkiksi ikä voidaan sovitaa joko jatkuvana muuttujana tai kategorisoituna (esim. alle 5-vuotiaat / 5–10-vuotiaat / yli 10-vuotiaat). Kun käytetään kovariaatteja sisältäviä tilastollisia malleja, on hyvä esittää vasteiden tulokset kovariaattien muodostamissa alaryhmissä, jotta mallin oletuksia voidaan arvioida. Herkkyyksianalyysit pitää suunnitella etukäteen ja esittää päävasteen analyysin luotettavuuden selvittämiseksi. Jos tulokset ovat ristiriitaisia, on tärkeä tuoda tämä pohdinnassa esiin. Vaikuttavuustutkimusten tulosten tulkinnan avulla lukijaa voidaan johtaa virheellisiin päätelmiin, jos tutkimusryhmien välisiä eroja ei voida loogisesti selittää, esimerkiksi kovariaateilla korjattujen ja korjaamattomien analyysien tulosten välillä. Ensisijaisen analyysin ei tulisi sisältää kovariaatin ja tutkimusryhmän välistä yhdysvaikutustermiä. Mikäli huomattavaa yhdysvaikutusta odotetaan etukäteen, tutkimus on hyvä suunnitella niin, että se mahdollistaa erilaiset arvot interventioiden vaikuttavuudesta kyseisen kovariaatin alaryhmissä.

Perustason kovariaattien käyttö on hyvin yleistä satunnaistettuja vertailukokeita raportoitaessa (Cognivia team, 2020). Vuonna 2019 julkaistun systemaattisen katsauksen mukaan neljässä tieteellisessä lehdessä vuosina 2009–2010 julkaistuissa satunnaisissa vertailututkimuksissa 84 prosentissa käytettiin kovariaatteja tilastollisissa malleissa. Näistä satunnaisista verailukokeista 91 prosentissa raportointiin, että kovariaatin sisällyttäminen tilastollisiin malleihin oli ennalta määritetty ennen analyysin suorittamista (Ciolino ym., 2019).

PUUTTUVAN TIEDON VAIKUTUS ANALYYSIIN

Puuttuva tieto on aina mahdollinen harhan lähde. Koeasetelmasta ja tutkimustyypistä riippumatta tutkimuksen suorittajien tavoitteena on kaikkien tutkittavien osalta tietojen täydellinen talteenotto, myös intervention lopettaneiden (European Medicines Agency, 1998). Satunnaistetuissa vertailukokeissa – kuten muissakin tutkimuksissa – kannattaa tehdä kaikki mahdollinen, jotta puuttuvalta tiedolta vältyttäisiin tai ainakin sen määrä voitaisiin minimoida. Vaikka käytännössä tämä ei toteutuisikaan, puuttuvia tietoja ei voida jättää huomioimatta analyyseissa ja niiden tulkinnassa. Tietojen puuttumiseen on monia mahdollisia syitä:

- tutkittava kieltäytyy jatkamasta tutkimusta
- tutkittava vetäytyy intervention epäonnistumisen vuoksi (*withdraw due to treatment failure*)
- tutkittava muuttaa toiselle paikkakunnalle
- tutkittava kokee intervention onnistuneen tai aiheuttaneen seurannais- tai liitännäisvaikutuksia, joista vain osa liittyy intervention.

Puuttuvaa tietoa voi olla monessa eri kohdassa: mittaukset voivat olla saatavilla pelkästään lähtötilanteessa, tai mittaukset voivat puuttua vain lähtötilanteesta tai ne voivat puuttua yhdestä, useammasta tai kaikista seuranta-arvioinneista. On myös mahdollista, että tutkittava suorittaa tutkimuksen loppuun, mutta osa tiedoista jää vain yksinkertaisesti ilmoittamatta tai keräämättä.

”Puuttuvia tietoja ei voida jättää huomioimatta analyyseissa ja niiden tulkinnassa.”

Satunnaistettu vertailukoe analysoidaan hoitoaioperiaatteen (*intention-to-treat* (ITT) *principle*) mukaisesti eli analyyseihin sisällytetään kaikki satunnaistetut tutkittavat (Moher ym., 2010; Hinkka-Yli-Salomäki, 2022a). Puuttuvan tiedon huomiotta jättäminen rikkoo tätä periaatetta (European Medicines Agency, 2011). Intervention vaikuttavuutta arvioitaessa on tärkeää sisällyttää kaikki tutkittavat analyysiin, eikä vain niitä, joilta onnistutaan keräämään kaikkien mittausajankohtien kaikki tieto (*per-protocol*-periaate). Toisin sanoen yleensä ei olla kiinnostuneita intervention vaikuttavuudesta sellaisessa epärealistisessa tilanteessa, jossa kaikki tutkittavat saavat heille satunnaistetun intervention ja seuraavat täydellisesti hoitosuunnitelmaa ja jossa tutkija saa kerättyä täydelliset tiedot vailla yhtäkään puuttuvaa tietoa.

Siihen, missä määrin puuttuvat arvot johtavat harhaisiin johtopäätöksiin intervention vaikutuksen suuruudesta, vaikuttavat monet tekijät. Näitä ovat esimerkiksi puuttuvan tiedon, satunnaistetun tutkimusryhmän ja vasteen välinen suhde; vasteen määrittelyyn käytetty menetelmä ja vasteiden odotettavissa olevat muutokset lähtötilanteesta seuranta-ajankohtiin. Valitettavasti menetelmä, se miten puuttuvia tietoja käsitellään, voi itsessään aiheuttaa harhaa. Puuttuvan tiedon hallintaan ei ole olemassa yhtä ainoaa oikeaa lähestymistapaa, joka sopisi kaikkiin tilanteisiin (European Medicines Agency, 2011).

Puuttuvan tiedon vaikutus tutkimuksen voimaan

Tutkimuksen voimaan eli siihen, miten hyvin tutkimus pystyy havaitsemaan tutkimusryhmien välisiä eroja, vaikuttaa otoskoko ja vasteen hajonta. Tutkimuksen voima kasvaa, jos otoskoko kasvatetaan tai vasteen hajonta pienenee. Jos tutkimusaineisto analysoidaan vain niiden tutkittavien osalta, joiden aineisto on täydellistä, otoskoko pienenee, koska analysoinnin ulkopuolelle jäävät ne tutkittavat, joiden aineisto on puutteellista. Tämä heikentää vasteen tilastollisen testin voimaa. Mitä enemmän puuttuvaa tietoa on, sitä todennäköisemmin voima pienenee. Tutkittavilla, joiden aineistosta puuttuu tietoa, saattaa todennäköisemmin olla vasteen ääriarvoja. Ääriarvoilla tarkoitetaan esimerkiksi jatkuvan muuttujan tapauksessa minimi- tai maksimiarvojen lähellä olevia arvoja tai Likert-asteikolla kerättävän kategorisen tiedon ääripäitä, joissa vaihtoehdot tavallisimmin ovat

täysin samaa mieltä tai *täysin eri mieltä*. Siksi näiden tutkittavien sulkeminen pois analyseista voi johtaa vasteen hajonnan aliarviomiseen ja siten keinotekoisesti kaventaa intervention vaikutuksen luottamusväliä. Jos puuttuvan tiedon käsittelyyn käytetyt menetelmät eivät huomioi riittävästi edellä mainittua tilannetta, ei koe- ja vertailuryhmien vasteen väliselle erolle laskettua luottamusväliä voida pitää luotettavana ja riittävänä kuvaamaan tutkittavan intervention vaikutuksen epävarmuutta.

”Mitä enemmän puuttuvaa tietoa on, sitä todennäköisemmin voima pienenee.”

Puuttuva tieto lisää tutkimuksen harhaa

Jos analyseista suljetaan pois tutkittavia, joiden aineistosta puuttuu tietoa, voi se vaikuttaa koe- ja vertailuryhmien vertailukelpoisuuteen ja tutkimusotoksen edustavuuteen (ns. ulkoiseen validiteettiin) suhteessa kohderyhmään. Puuttuvien tietojen ei odoteta aiheuttavan harhaa, jos ne eivät liity havaitsemattoman vasteen todelliseen arvoon. Toisin sanoen vasteen huonojen arvojen tulisi puuttua yhtä todennäköisesti kuin hyvien arvojen. Jos puuttuva tieto liittyy vasteen todelliseen arvoon, se voi aiheuttaa harhaa intervention vaikutuksen arvioon, vaikka puuttuva tieto ei liittyisikään tutkittavaan interventioon (eli puuttuva tieto on yhtä todennäköistä kaikissa tutkimusryhmissä, myös vertailuryhmissä). Mikäli puuttuva tieto riippuu sekä tutkittavasta interventtiosta että vasteen arvoista, voi puuttuvan tiedon käsittelemättä jättäminen aiheuttaa harhaa. Esimerkiksi puuttuva tieto voi olla todennäköisempää vertailuryhmässä, jossa annettu vertailuinterventio voi olla tehottomampi kuin koeryhmässä annettu interventio. Toisin sanoen vertailuryhmän tutkittavat voivat keskeyttää intervention useammin, koska kokevat sen hyödyttömämmäksi kuin koeryhmän tutkittavat.

Menetelmiä puuttuvan tiedon käsittelemiselle

Satunnaistetun vertailukokeen tulosten tulkinta on erityisen ongelmallista, jos puuttuvaa tietoa on paljon. Valitettavasti ei ole olemassa yleisesti hyväksyttyä nyrkkisääntöä puuttuvan tiedon enimmäismäärälle. Tavallisesti tutkijan arvioon puuttuvasta tiedosta vaikuttavat esimerkiksi vasteen luonne, tutkimuksen kesto, interventioon sitoutuminen (tutkimushenkilöiden sitoutuminen tutkimussuunnitelmaan on psykiatrian alalla perinteisesti ollut heikompaa kuin muilla aloilla). Satunnaistettua vertailukoetta voidaan pitää laadukkaana ja luotettavana, jos menetelmät puuttuvan tiedon käsittelemiseen ovat järkeviä ja etukäteen tutkimussuunnitelmassa määritellyjä (European Medicines Agency, 2011). Menetelmien määrittelyä voidaan muuttaa ja tarkentaa tilastolliseen analyysisuunnitelmaan vielä päävasteen analyysin aikana, jos tutkimuksen analysoinnista vastaava henkilö on tällöin vielä sokkona (eli hänellä ei ole tiedossa, mihin tutkimusryhmään tutkittava kuuluu).

”Satunnaistettua vertailukoetta voidaan pitää laadukkaana ja luotettavana, jos menetelmät puuttuvan tiedon käsittelemiseen ovat järkeviä ja etukäteen tutkimussuunnitelmassa määritellyjä.”

Vaikka yhtä ainoaa oikeaa tapaa käsitellä puuttuvaa tietoa ei ole olemassa, on olemassa yleisesti hyväksytyjä ohjeita puuttuvan tiedon käsittelemiseen. Jotta vältetään aineistolähtöiseltä valinnalta puuttuvan tiedon käsittelemiseksi, on välttämätöntä määrittää valitut menetelmät etukäteen tutkimussuunnitelmassa ja se, mitä menetelmää käytetään ensisijaisessa analyysissä. Yleensä mitä pidempi seuranta-aika on, sitä suurempi on puuttuvan tiedon todennäköisyys.

TAULUKKO 1.

Imputointimenetelmiä.

Imputointimenetelmä	Sopivuus
Viimeisen havainnon jatkaminen eteenpäin (<i>last observation carried forward, LOCF</i>)	<ul style="list-style-type: none"> Kun useilta tutkittavilta useita perustason mittauksen jälkeisiä toistomittauksia. Kun tutkittava esimerkiksi keskeyttää ennen tutkimuksen loppumista, niin hänen viimeistä mitattua vasteen arvoaan käytetään kaikissa myöhemmissä aikapisteissä. Voi tuottaa hyvinkin harhaisia päävasteiden ja toissijaisien vasteiden estimaatteja; tilanteesta riippuen joko tutkittavan intervention vaikutukselle liian konservatiivisia tai ylioptimistisia estimaatteja.
Perustason havainnon jatkaminen eteenpäin (<i>baseline observation carried forward, BOCF</i>)	<ul style="list-style-type: none"> Kun tutkittavilla ei ole perustason jälkeisiä intervention aikaisia tai sen jälkeisiä mittauksia. Perustuu ajatukseen, että tutkimushenkilön vaste pysyy alkutilanteessa, mikäli kyse on pitempiaikaisesta sairaudesta, kuten masennuksesta.
Parhaimman tapauksen imputointi (<i>best case imputation</i>)	<ul style="list-style-type: none"> Sopii positiivisille keskeyttäneille tai pudonneille eli toisin sanoen niille, joilla interventio on toiminut täydellisesti (ns. parantuneet).
Huonoimman tapauksen imputointi (<i>worst case imputation</i>)	<ul style="list-style-type: none"> Sopii negatiivisen syyn takia keskeyttäneille eli niille, joille interventio ei ole toiminut.

Puuttuvan tiedon huomioivan menetelmän valintaan vaikuttavat

- tutkittavien vetäytymiseen liittyvät erot tutkimusryhmien välillä
- vetäytymisten ajankohtaan liittyvät erot tutkimusryhmien välillä
- tutkittavien vetäytymisten syyt.

Täydellisen tiedon analyysia (*complete case analysis*) ei suositella ensisijaisena analyysinä vaikuttavuustutkimuksissa. Menetelmässä puuttuva tieto jätetään huomioimatta, ja vain tutkimushenkilöt, joilta ei puutu tietoa, otetaan mukaan analyyseihin. Seurauksena on hoitoaieperiaatteen noudattamatta jättäminen sekä alttius harhaan (European Medicines Agency, 2011). Herkkyystarkasteluissa täydellisen tiedon analyysit ovat sen sijaan suositeltavia tulosten robustisuuden havainnollistamiseksi.

Yleisimmin käytetyt tilastolliset menetelmät puuttuvan tiedon huomioimisessa ovat

- 1) imputointimenetelmät, joilla paikataan puuttuvaa tietoa
- 2) sekamallit (*mixed-effect models for repeated measures*) jatkuville vasteille ja yleistetyt estimointiyhtälöt (*generalized estimating equations*) kategorisille vasteille ja lukumäärävasteille
- 3) elinaika-analyysit (survival analysis).

Imputointimenetelmät

Yksittäisten puuttuvien arvojen imputoinnissa puuttuvat tiedot korvataan yksittäisillä arvoilla ja analyysi suoritetaan kuin puuttuvia tietoja ei olisi ollenkaan. Taulukossa 1 on esimerkkejä tavallisimmista imputointimenetelmistä.

Uusi, täysin hyväksyttävä tapa on soveltaa eri imputointimenetelmiä, kuten LOCF- ja BOCF-menetelmiä, eri syiden takia puuttuviin havaintoihin. Tämä edellyttää menetelmien määrittelyä etukäteen tutkimussuunnitelmassa. Yksittäisten puuttuvien arvojen imputointi voi johtaa intervention vaikutuksen yliarviointiin eli todellisuutta pienempään vaihteluun (European Medicines Agency, 2011). Moninkertaisessa imputoinnissa (*multiple imputation*) alkuperäisestä aineistosta luodaan useita kopioita, joissa puuttuvat tiedot on korvattu jonkin sopivan stokastisen eli sattuman huomioivan mallin avulla saaduilla arvoilla. Ideana on, että useat (jopa tuhannet) moninkertaisen imputoinnin avulla saadut kopioaineistot analysoidaan täydellisinä aineistoina. Tämän jälkeen kopioaineistoista saadut koe- ja vertailuryhmien vasteiden estimaatit eli lasketut arvot yhdistetään lopuksi yhdeksi piste-estimaatiksi.

Sekamallit ja yleistetyt estimointiyhtälöt

Sekamallien käytöllä vältytään puuttuvien tietojen imputoinnilta. Sekamallit sopivat tilanteisiin, joissa havainnot ovat korreloituneita, kuten tavallisesti useita toistomittauksia sisältävissä satunnaisissa vertailukokeissa. Kun kyse on jatkuvasta vasteesta, lineaariset

sekamallit ovat toistomittaustilanteessa (*mixed model for repeated measures*, MMRM) erinomainen vaihtoehto, koska niissä puuttuvien tietojen imputoinnin sijaan käytetään olemassa olevaa aineistoa pohjana uskottavuusprofiilien muodostamisessa kaikille havainnoille – myös puuttuville. MMRM käyttää tavallisesti rajoitettua suurimman uskottavuuden estimointimenetelmää (*restricted maximum likelihood tai residual maximum likelihood*, REML). Oleellista lineaarisen sekamallin muodostamisessa on aineistoon sopivimman kovarianssirakenteen valinta. REML-estimointimenetelmän avulla muodostuu vasteprofiileja, jolloin kaikkia havaittuja arvoja voidaan käyttää mukana analyyseissä. Kategorisille vasteille ja lukumäärävasteille vastaavana menetelmänä voidaan käyttää yleistyttä estimointiyhtälöitä.

Elinaika-analyysi

Elinaika-analyysien käyttäminen on tavallista, kun vasteena on aika aloituksesta päätetapahtumaan (*time-to-event*). Analyyseissä huomioidaan sensuroidut havainnot, jotka syntyvät tilanteissa, joissa tutkittavan seuranta päättyy esimerkiksi muuton takia tai päätetapahtuma (*event*) ei tapahdu seuranta-aikana. Esimerkiksi tutkittaessa uuden, nuorten nopeampaan työllistymiseen tähtäävän intervention vaikuttavuutta, voi olla mahdollista, että kaikki tutkittavat eivät työllisty seuranta-aikana. Tällöin puhutaan sensuroidusta havainnosta, koska tutkittavan seuranta päättyy seuranta-ajan päättymiseen, ei päävasteeseen eli työllistymiseen. Toisin sanoen kaikille ei tapahdu toivottua päätetapahtumaa, kun vasteena on aika intervention aloittamisesta työllistymiseen.

Herkkyyssanalyytit puuttuvaa tietoa sisältävälle aineistolle

Herkkyyssanalyytit (*sensitivity analyses*) ovat tärkeitä, jos puuttuvaa tietoa on runsaasti. Puuttuvan tiedon ollessa vähäistä herkkyyssanalyytien tarkoituksena on lähinnä tukea ensisijaisia analyysejä. Yksityiskohtaista ohjetta herkkyyssanalyytien sisällöstä on vaikea antaa, mutta

- ITT-periaatteen mukaisesti tehtyjen analyyseiden tuloksia kannattaa verrata täydelle aineistolle tehtyjen analyyseiden (per protocol) tuloksiin
- mikäli puuttuvan tiedon määrä on merkittävän suuri, kannattaa verrata eri menetelmillä tehtyjen analyyseiden tuloksia keskenään, esimerkiksi imputoidun aineiston ja sekamallien avulla saatuja tuloksia
- tuloksia voi verrata ns. responder-analyysin¹ tuloksiin, jos vaste on jatkuva
- tuloksia voi verrata huonoimman tapauksen analyysin tuloksiin (*worst case analysis*)².

¹ Responder-analyysissa ennalta määritetyn katkaisupisteen (*cut-off point*) mukaan tutkittavat luokitellaan kahteen luokkaan: interventio onnistunut ja interventio epäonnistunut. Ensisijaisia analyysejä tukevista responder-analyyseissa tutkittavat, joilta puuttuu havainnot – tai vaihtoehtoisesti tietyistä syistä puuttuu havainnot – luokitellaan interventio epäonnistunut -luokkaan. Responder-analyyseissa vaste on kategorinen ja analysoidaan kuten kategoriset muuttujat.

² Huonoimman tapauksen analyysissa vertailuryhmän puuttuvat havainnot korvataan parhailla mahdollisilla vasteen tuloksissa. Vastaavasti koeryhmän puuttuvat tulokset korvataan huonoimmilla mahdollisilla vasteen arvoilla. Jos tämä äärimmäinen analyysi osoittaa tutkittavan interventioyöryhmän tulokset tilastollisesti ja kliinisesti merkitseviksi vertailuryhmään verrattuna, voidaan ensisijaisen analyyseiden tuloksia pitää luotettavina puuttuvista tiedoista huolimatta.

Tutkimuksen laadun varmentamiseksi herkkyysanalyysit on tärkeää suunnitella etukäteen ja kirjata tutkimussuunnitelmaan tai viimeistään tilastolliseen analyysisuunnitelmaan ennen satunnais-
tamisen koodin avaamista (European Medicines Agency, 1998).

MONIVERTAILU ANALYSEISSA

Satunnaistetuissa vertailukokeissa on usein välttämätöntä saada vastaus useampaan kysymykseen samanaikaisesti. Tämä johtaa monivertailuihin, joissa tehdään samanaikaisesti useita tilastollisia testejä. Kun useita tilastollisia testejä tehdään samalla aineistolla, niin väärän positiivisen (false positive) sattumalöydöksen mahdollisuus kasvaa – eli sattumalta löydetäänkin tilastollisesti merkitsevä tulos, vaikka näin ei todellisuudessa olisi. Toisin sanoen nollahypoteesin hylkäämisen todennäköisyys kasvaa. Virheen todennäköisyys kasvaa testien lukumäärän mukaan seuraavasti:

P (vähintään yksi merkitsevä tulos) = $1 - (1 - \alpha)^k$, missä α on merkitsevyytaso ja k on testien lukumäärä.

Jos samanaikaisesti tehdään kaksi testiä 5 prosentin merkitsevyytastolla ($\alpha = 0.05$), todennäköisyys tehdä vähintään yksi virheellinen löydös on yli 9 prosenttia. Kolmen testin tapauksessa virhetodennäköisyys on jo yli 14 prosenttia. Monivertailuun perustuvaa päättelyä tehdään kaikissa vertailukokeissa, myös satunnais-
tetuissa, koska esimerkiksi vasteita ja mittausajankohtia on yleensä enemmän kuin yksi. Jopa vertailuryhmiä voi olla enemmän kuin kaksi, mikä moninkertaistaa tehtyjen parittaisten vertailujen lukumäärän. Yleisimmät monivertailuongelmaan liittyvät tilanteet ovat seuraavat:

- päävasteita on enemmän kuin yksi
- ryhmien välisiä vertailuja on enemmän kuin yksi
- testataan toistuvasti eri aikapisteissä
- tehdään vähintään yksi välianalyysi (*interim analysis*)³.

Milloin monivertailua ei tarvitse huomioida?

Monivertailua ei tarvitse huomioida satunnaistetussa vertailukokeessa, jossa on kaksi ryhmää: koe- ja vertailuryhmä, yksi päävaste, yksi päävasteeseen liittyvä tutkimushypoteesi sekä etukäteen määritelty analyysisuunnitelma, ja jossa ei ole suunniteltu tehtäväksi välianalyysijä. Edellä mainittujen tilanteiden lisäksi on myös muita tilanteita, joissa monivertailua ei tarvitse huomioida. Esimerkiksi silloin, kun toissijaisten vasteiden analyysit on määritelty ainoastaan päävasteita tukeviksi analyyseiksi (European Medicines Agency, 2002). Edellä mainitussa tilanteessa tarkoituksena on osoittaa kaikkien päävasteiden tilastollinen merkitsevyys, kun ryhmiä vertaillaan keskenään.

Monivertailuongelman välttämiseksi satunnaistetuissa vertailukokeissa kannattaa käyttää vain yhtä päävastetta (European Medicines Agency, 1998). Yksi päävaste riittää, jos on yleinen yksimielisyys siitä, että intervention aiheuttama muutos tässä tiettyssä vasteessa riittää osoittamaan yksinään intervention vaikuttavuuden. Jos näin ei ole, tarvitaan useampi päävaste. Monivertailun huomioimista ei tarvitse, mikäli tutkimussuunnitelmassa on etukäteen määritelty, että kaikissa ensisijaisissa analyyseissa tilastollisen merkitsevyyden osoittaminen on välttämätöntä. Kahden tai useamman

päävasteen tilanteessa tutkimus voidaan suunnitella niin, että päävasteet järjestetään kliinisen merkityksen mukaiseen järjestykseen. Tällöin monivertailua ei tarvitse huomioida. Hierarkkinen järjestys tutkimushypoteesien testauksessa on määriteltävä tarkasti etukäteen tutkimussuunnitelmassa (European Medicines Agency, 2002).

”Monivertailuongelman välttämiseksi satunnaistetuissa vertailukokeissa kannattaa käyttää vain yhtä päävastetta.”

Esimerkkinä tästä on tutkimus, jossa ensimmäinen hypoteesi liittyy siihen, miten vaikuttava interventio on vähentämään tutkittavien raskauden jälkeisiä masennusoireita puolen vuoden päästä lapsen syntymästä. Toinen hierarkkisesti alempana oleva hypoteesi liittyy siihen, miten interventio vaikuttaa tutkittavien terveyskeskuskäynteihin minkä tahansa mielialaoireiden vuoksi kaksi vuotta lapsen syntymän jälkeen. Mikäli ensimmäiseen hypoteesiin liittyvä tilastollisen testin tulos ei ole merkitsevä, mutta toiseen hypoteesiin liittyvä testin tulos sen sijaan on, intervention ei voida todeta olevan vaikuttava verrattuna vertailuryhmässä annettuun interventioon.

Toisinaan tutkimuksissa ollaan kiinnostuneita tekemään päävasteen analyysejä eri tutkittavien osajoukoille. Mikäli satunnaistetun vertailukokeen laatu halutaan varmentaa, on osajoukkoanalyysit suunniteltava etukäteen ja määriteltävä tutkimussuunnitelmassa. Mikäli osajoukkoanalyysijä käytetään tukemaan pääanalyyseiden tuloksia, ei monivertailun huomioimista tarvita.

Erilaisia tilastollisia malleja tai menetelmiä kokeillaan toisinaan samaan aineistoon (esimerkiksi parametrisen testi (Wilcoxonin testi) vs. epäparametrisen testi (log-rank testi)). Kokeilun perusteella tehty analyysistrategian valinta aiheuttaa monivertailuongelman sekä harhaa, erityisesti jos tutkimusta analysoidaan on tieto tutkittavien saamista interventioista. Tämän vuoksi päävaste on hyvä analysoida sokkona ja valita analyysistrategia ennen satunnaistamiskoodin avaamista. (Hinkka-Yli-Salomäki, 2021).

Usein satunnaistetuissa vertailukokeissa käytetään yhdistelmämuuttujia (*composite variables*) esimerkiksi erilaisten psykiatristen ja neurologisten häiriöiden tunnistamiseksi. Näissä kyselyjen avulla luodaan yhdistelmämuuttujia laskemalla yksittäisten kysymysten vastausten pistemääriä yhteen yleisesti hyväksytyyn algoritmin avulla. Toisin sanoen, jos esimerkiksi lapsen käyttöhäiriötä tutkivassa kyselyssä on viisi kysymystä, ja kussakin yksittäisessä kysymyksessä on kolme vastausvaihtoehtoa (0 = ei totta, 1 = jokseenkin totta, 2 = täysin totta), saadaan käyttöhäiriön yhdistelmämuuttujia tutkittaville laskemalla annetut pisteet yhteen (vaihteluväli 0–10). Toisinaan yhdistelmämuuttujia voi olla yhdessä kyselyssä useita.

Esimerkkinä tästä on Vahvuudet ja vaikeudet -kysely (*Strengths and Difficulties Questionnaire*, SDQ), jossa 25 kysymyksestä johdetaan usein kokonaispistemäärämuuttujan lisäksi viisi summa-

³ Välianalyysi on käynnissä olevan tutkimuksen arviointi. Välianalyysissa tehdään ensisijainen analyysi joko osalle vasteita tai kaikille. Välianalyysin tuloksilla voi olla mahdollisesti vaikutusta tutkimuksen kulkuun. Yleisimmin välianalyysijä tehdään suurissa ja pitkäkestoisissa tutkimuksissa, joissa halutaan esimerkiksi selvittää mahdollisuutta päättää tutkimus aiottua nopeammin tai varmistua uuden intervention turvallisuudesta.

muuttujaa, jotka liittyvät tunne-elämään, kaverisuhteisiin, prososiaalisuuteen, yliaktiivisuuteen ja käytökseen (*emotional, peer problems, prosocial behaviour, hyperactive and conduct scores*). Näin erilaisten summamuuttujien analysoinnissa ei monivertailua tarvitse huomioida. Jos yhtä kyselystä johdettua yhdistelmämuuttujaa käytetään tutkimuksen päävasteena, voidaan samasta kyselystä johdettuja muita yhdistelmämuuttujia käyttää toissijaisina vasteina. Tutkimussuunnitelmaan tulee liittää kyselyn psykometrisia ominaisuuksia, kuten luotettavuus ja validiteetti. Näiden analysoinnissa ja raportoinnissa ei tarvitse huomioida monivertailua.

”Usein satunnaistetuissa vertailukokeissa käytetään yhdistelmämuuttujia (composite variables) esimerkiksi erilaisten psykiatristen ja neurologisten häiriöiden tunnistamiseksi.”

Tavallisimpia monivertailun huomioivia menetelmiä

Monivertailu voidaan huomioida usealla eri tavalla niin, että merkitsevyytensä voidaan pitää kokonaisuutena hallinnassa. Tähän on kehitetty monia tilastollisia menetelmiä, jotka kontrolloivat virheellisen hylkäämisen todennäköisyyttä. Käytännössä yksittäisten testien merkitsevyytensä muokataan erilaisten algoritmien mukaan. Konservatiivinen, analyysien voimaa vähentävä, mutta teknisesti helpoin tapa on ns. Bonferroni-korjaus, jossa tavoiteltu merkitsevyytensä jaetaan testien lukumäärällä, ja näin laskettua arvoa käytetään jokaisessa yksittäisessä testissä. Kun k on esimerkiksi koe- ja vertailuryhmien tai tutkittavalta toistettavien mittausten lukumäärä ja tehdään kaikki parittaiset vertailut, niin vertailujen lukumäärä $m = k \cdot (k - 1) / 2$. Toisin sanoen, jos koe- ja vertailuryhmiä on kolme, saadaan vertailujen lukumääräksi kolme. Bonferroni-epäyhtälöön $\alpha \leq m \cdot \alpha^*$ perustuva monivertailuongelman huomioiva korjaus on $\alpha^* = \alpha / m$, jossa α^* on uusi merkitsevyytensä. Toisin sanoen, jos vertailujen lukumäärä on esimerkiksi 3 ja yhden vertailuun liittyvän tilastollisen testin p -arvoksi saadaan 0,03, on Bonferroni-korjattu p -arvo $3 \cdot 0,03 = 0,09$. Tässä esimerkissä siis Bonferroni-korjamaton p -arvo 0,03 on tilastollisesti merkitsevä, kun taas Bonferroni-korjattu p -arvo 0,09 on ei-merkitsevä. Toisin sanoen, yksittäisten testien p -arvot kerrotaan vertailujen lukumäärällä, ja saatuja vakioituja (*adjusted*) p -arvoja verrataan merkitsevyytensä 0,05. Vähemmän konservatiivisia, ja samalla monimutkaisempia tapoja ovat

- Šidák-korjaus
- Benjamini-Hochberg-korjaus
- Dunnett-korjaus
- hylkäysvirhemenetelmät.

Kun tutkitaan suuria aineistoja, joissa voidaan tehdä kymmeniä, satoja monivertailuja (esimerkiksi geenitutkimuksissa), Bonferroni-menetelmä on liian konservatiivinen eli kerrottaessa saadut p -arvot vertailujen lukumäärällä saadaan harvemmin merkitseviä tuloksia, minkä vuoksi suositetaan on kasvattanut hylkäysvirhemenetelmää (*False-discovery rate, FDR*). Hylkäysvirhemenetelmään liittyvä hylkäysvirheaste on odotettu virheellisten löydösten eli nollahypoteesia tukevien testien osuus kaikkien hylättyjen nollahypoteesien joukossa. Esimerkkinä ovat yhdistelmä tutkimukset,

joissa tutkittavilta kerätään paljon tietoa satunnaistetuissa vertailukokeissa kyselyjen, rekisterien ja biologisten näytteiden avulla. Tämä voi johtaa satoihin jopa tuhansiin samanaikaisiin tilastollisiin testeihin. Glickman kollegoineen (2014) suosittelee hylkäysvirhemenetelmän käyttämistä perinteisen p -arvojen korjausten sijaan. Toistaiseksi p -arvojen korjaaminen on suosituin tapa, vaikka sen tuottamat tulokset ovat usein liian konservatiivisia. Hylkäysvirhemenetelmän käyttö on toistaiseksi vierasta, koska menetelmän käyttö vaatii erillistä ohjelmointia eli sitä ei ole valmiina yleisimmissä perustason tilasto-ohjelmissa. Otskoon kasvattaminen ei vaikuta hylkäysvirhemenetelmän toimivuuteen. P -arvojen korjausmenetelmissä otskoon kasvattaminen vaikuttaa yleisesti konservatiivisesti hylkäyspäätöksiin, mutta hylkäysvirhemenetelmän on todettu toimivan hyvin sekä isoilla että pienillä otoskoilla (Glickman ym., 2014).

”Hylkäysvirhemenetelmän on todettu toimivan hyvin sekä isoilla että pienillä otoskoilla.”

OSAJOUKKOANALYYSIT SATUNNAISTETUISSA VERTAILUKOKEISSA

Satunnaistetut vertailukokeet keskittyvät yleensä tutkimussuunnitelmassa ennalta määritettyjen vasteiden eron tarkasteluun koe- ja vertailuryhmien välillä. Tarkoituksena on selvittää, onko tarkastelun kohteena oleva interventio yleisesti hyödyllinen. Kuitenkin jotkut interventiot voivat toimia eri tavoin erilaisilla osajoukoilla. Jotta interventio voidaan tarjota mahdollisimman yksilöllisesti, tulisi tutkimuksen avulla selvittää ne osajoukot, jotka todennäköisimmin hyötyvät interventiosta ja ne, jotka eivät todennäköisimmin hyödy siitä lainkaan tai hyvin vähän. Osajoukkoanalyysiin perustuva raportointi on usein harhaanjohtavaa pienen otoskoon ja monivertailuongelman vuoksi (Wang ym., 2007). Osajoukot voidaan määrittää seuraavien muuttujien perusteella:

- dikotominen (esim. sukupuoli: nainen vs. mies)
- kategorinen (esim. asuinympäristö: kaupunki vs. taajama vs. maaseutu)
- järjestysasteikollinen (esim. taudin perustason vaikeusaste)
- jatkuva (esim. ikä).

Osajoukot voidaan määrittää myös eri muuttujista koostuvien yhdistelmämuuttujien avulla, joita kutsutaan usein riskipistemääräksi (*risk scores*).

Yleensä satunnaistamista ei suunnitella erilaisille osajoukoille. Näin osajoukot, joille analyysi tehdään, voivat olla perustasolla mitattavien muuttujien suhteen epätasapainossa. Intervention vaikutavuuden osalta tulee selvittää, että vaikuttavuus ei johdu osajoukkojen perustason epätasapainosta. Useita osajoukkoanalyysia tehtäessä monivertailu on ehdottomasti huomioitava. Yhdysvaikutusten testaamista tilastollisilla malleilla pidetään hyväksyttävänä tapana päättää osajoukkoanalyysin tarpeesta. Koska testit ovat usein heikkovoimaisia pienen otoskokonsa vuoksi, päätös kannattaa perustaa sekä tilastotieteeseen että kliiniseen arviointiin. Julkaisuissa pelkkää yhdysvaikutustermin p -arvon raportointia ei pidetä riittävänä päätöksenteon perustana. On tärkeää esittää estimaatit luottamusväleinen ja arvioida intervention vaikutuksia osajoukoissa sekä havaittujen erojen kliinistä merkitystä. Monet tieteelliset lehdet suosivat metsikkökuviota (*forest plot*), jossa esitetään intervention vaikutukset kaikille tehdyille osajoukoille piste-estimaatteineen ja

luottamusväleinen. Kuvasta lukija näkee nopeasti esimerkiksi mahdollisen pienen otoskoon leveän luottamusvälin todennäköisenä syynä, jolloin kyseisen osajoukon tulosta ei voida pitää välttämättä kovin luotettavana (European Medicines Agency, 2019).

”Monet tieteelliset lehdet suosivat metsikkökuviota (forest plot), jossa esitetään intervention vaikutukset kaikille tehdyille osajoukoille piste-estimaatteineen ja luottamusväleineen.”

Satunnaistettujen vertailukokeiden osajoukkoanalyysit on suurimmaksi osaksi (94,9 %) tehty kategorisoimalla jatkuva muuttuja (Williamson ym., 2022). Monimutkaisimpien menetelmien käyttö osajoukkojen muodostamisessa, jossa jatkuva muuttuja pidettäisiin jatkuvana, on käytännössä hyvin harvinaista. Tämä on harmillista, koska tiedetään kuinka jatkuvan muuttujan luokittelu, tavallisesti dikotomisointi kahteen osajoukkoon, johtaa informaation menetykseen ja sitä kautta tilastollisen testin heikompaan voimaan. Toisaalta myös virheellisten positiivisten tulosten riski voi kasvaa. Austin ja Brunner tutkivat, että riski kasvaa otoskoon kasvaessa, koska riskitekijän ja kovariaatin välinen korrelaatio tavallisesti kasvaa. Ilmiö on sitä selvemmin havaittavissa, mitä vähemmän muuttujalla on luokkia (Austin & Brunner, 2004). Kun jatkuva muuttuja luokitellaan, niin katkaisupisteen valinta on hyvä perustella julkaisussa (Williamson ym., 2022).

YHTEENVETO

Ensisijaisen analyysin tulisi sisältää vain tutkimussuunnitelmassa etukäteen määritellyt kovariaatit, ei käsittelyryhmän ja kovariaatin yhdysvaikutustermiä. Tämän lisäksi suosituksena on, että käsittelyryhmän ja kovariaattien yhdysvaikutuksia tutkitaan. Nämä analyysit ovat kuitenkin luonteeltaan eksploratiivisia, ja niiden tehtävänä on lähinnä arvioida ensisijaisen analyysin tulosten robustisuutta.

Satunnaistetuissa vertailukokeissa tavoitteena on saada kerättyä täydellinen aineisto kaikilta tutkittavilta. Todellisuudessa on kuitenkin vaikea välttää puuttuvaa tietoa. Sen huomiotta jättäminen ei ole hyväksyttävää, kun suunnitellaan, toteutetaan ja raportoidaan tutkimusta. Toisin sanoen, satunnaistetuissa vertailukokeissa puuttuvan tiedon huomiotta jättäminen rikkoo hoitoaieperiaatetta (*intention-to-treat* (ITT) *principle*).

Monivertailuongelman välttämiseksi satunnaistetuissa vertailukokeissa kannattaa käyttää vain yhtä päävastetta. Mikäli päävasteita on useita, niin monivertailun huomioimista ei tarvitse, mikäli tutkimussuunnitelmassa on etukäteen määritetty, että kaikissa ensisijaisissa analyyseissa tilastollisen merkitsevyyden osoittaminen on välttämätöntä. Monivertailun huomioimista ei myöskään tarvitse, mikäli päävasteet järjestetään kliinisen merkityksen mukaiseen järjestykseen olettaen, että hierarkkinen järjestys tutkimushypoteesien testauksessa on määritetty tutkimussuunnitelmassa etukäteen.

Toisinaan interventiot voivat toimia eri tavoin erilaisilla osajoukoilla. Osajoukkoanalyysien tulokset on tärkeä esittää estimaattien ja luottamusvälien avulla, sillä pelkkä käsittelyryhmän ja kovariaatin yhdysvaikutuksen *p*-arvon esittäminen ei ole riittävää.

Kovariaatit, puuttuvan tiedon käsittelymenetelmät, herkkyysanalyysit, osajoukkoanalyysit ja tilastollisen hypoteesien järjestys on määrittävä ennakkoon tutkimussuunnitelmassa tai viimeistään ennen satunnaistamiskoodin avaamista. ●

AVAINSANAT:

satunnaistettu vertailukoe, perustaso, kovariaatti, puuttuva tieto, monivertailu, osajoukkoanalyysit, raportointi, artikkeli, laatu

SIDONNAISUUDET:

Sanna Hinkka-Yli-Salomäki: Ei sidonnaisuuksia.

LÄHTEET:

Austin, P. C. & Brunner, L. J. (2004). Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. *Stat Med*, 23(7):1159-78. <https://doi.org/10.1002/sim.1687>

Ciolino, J. D., Palac, H. L., Yang, A., Vaca, M. & Belli, H. M. (2019). Ideal vs. real: A systematic review on handling covariates in randomized controlled trials. *BMC Med Res Methodol*. 19(1):1-11. <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-019-0787-8>

Clifton, L. & Clifton, D. A. (2019). The correlation between baseline score and post-intervention score, and its implications for statistical analysis. *Trials* 20, 43. <https://doi.org/10.1186/s13063-018-3108-3>

Cognivia team (2020). Regulatory Guidelines for the Use of Baseline Covariates to Increase Clinical Trial Study Power. Haettu 1.3.2023 osoitteesta <https://cognivia.com/regulatory-guidelines-for-the-use-of-baseline-covariates-to-increase-clinical-trial-study-power/>

Egbewale, B. E. (2015). Statistical issues in randomised controlled trials: a narrative synthesis. *Asian Pac J Trop Biomed*, 5(5): 354-359. [https://doi.org/10.1016/S2221-1691\(15\)30367-1](https://doi.org/10.1016/S2221-1691(15)30367-1)

LÄHTEET:

- European Medicines Agency. (1998). ICH E9 statistical principles for clinical trials. ICH Harmonised Tripartite Guideline. <https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline>
- European Medicines Agency. (2002). Points to consider on multiplicity issues in clinical trials. <https://www.ema.europa.eu/en/multiplicity-issues-clinical-trials-scientific-guideline>
- European Medicines Agency. (2011). Guideline on missing data in confirmatory clinical trials. <https://www.ema.europa.eu/en/missing-data-confirmatory-clinical-trials-scientific-guideline>
- European Medicines Agency. (2015). Guideline on adjustment for baseline covariates in clinical trials. <https://www.ema.europa.eu/en/adjustment-baseline-covariates-clinical-trials-scientific-guideline>
- European Medicines Agency. (2019). Guideline on the investigation of subgroups in confirmatory clinical trials. <https://www.ema.europa.eu/en/investigation-subgroups-confirmatory-clinical-trials-scientific-guideline>
- Glickman, M., Rao, S & Schultz, M. (2014). False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *Journal of Clinical Epidemiology*, 67(8):850-857. <https://doi.org/10.1016/j.jclinepi.2014.03.012>
- Hinkka-Yli-Salomäki, S. (2021). Vaikuttavuustutkimus: Satunnaistettu vertailukoe tulee suunnitella huolella. *Kasvun tuki*. Haettu 1.3.2023 osoitteesta <https://kasvuntuki.fi/aikakauslehti/julkaisut/kasvun-tuki-aikakauslehti-2021/hinkka-yli-salomaki/>
- Hinkka-Yli-Salomäki, S. (2022a). Vaikuttavuustutkimus: Satunnaistetun vertailukokeen laadukkaan raportoinnin edellytykset. *Kasvun tuki*. Haettu 1.3.2023 osoitteesta <https://kasvuntuki.fi/aikakauslehti/julkaisut/kasvun-tuki-aikakauslehti-2022/hinkka-yli-salomaki/>
- Hinkka-Yli-Salomäki, S. (2022b). Vaikuttavuustutkimus: Miten arvioida harhaa satunnaistetuissa vertailukokeissa?. *Kasvun tuki*. Haettu 1.3.2023 osoitteesta <https://kasvuntuki.fi/aikakauslehti/julkaisut/kasvun-tuki-aikakauslehti-2022/hinkka-yli-salomaki/>
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux P. J., ... & Altman D. G. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*, 340:c869. <https://doi.org/10.1136/bmj.c869>
- Vickers, A. J. & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *BMJ*; 323:1123. <https://doi.org/10.1136/bmj.323.7321.1123>
- Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J. & Drazen, J. M. (2007). Statistics in medicine - reporting of subgroup analyses in clinical trials. *N Engl J Med*, 357:2189-2194. <https://doi.org/10.1056/NEJMSr077003>
- Williamson, S. F., Grayling, M. J., Mander, A. P., Noor, N. M., Savage, J. S., Yap, C. & Wason, J. M. S. (2022). Subgroup analyses in randomized controlled trials frequently categorized continuous subgroup information. *Journal of Clinical Epidemiology*, 150:72-79. <https://doi.org/10.1016/j.jclinepi.2022.06.017>