

## Vaikuttavuustutkimus:

# Satunnaistettu vertailukoe tulee suunnitella huolella

- Vaikuttavuustutkimuksessa pyritään osoittamaan mahdollisimman yksiselitteisesti, että saatu vaikutus on kokeessa riippumattomaksi muuttujaksi määritellyn tekijän tuottama.
- Satunnaistettua vertailukoetta pidetään luotettavimpana tapana mitata intervention vaikuttavuutta.
- Satunnaistamisella pyritään siihen, että ryhmät ovat mahdollisimman samankaltaisia muiden kuin tutkittavien vaikuttavien tekijöiden osalta.
- Koetta suunniteltaessa tulee määritellä kohdejoukko, verrokkiryhmät, päämuuttuja ja mahdolliset toissijaiset vasteet sekä nollahypoteesi. Myös otoskoko tulee laskea.
- Suunnittelussa tarvitaan tietoa tutkimuslainsäädännöstä, lupaprosessista, eettisistä asioista, statistiikasta ja tutkimusdatan keräämisestä.



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

Vaikuttavuustutkimuksessa pyritään osoittamaan mahdollisimman yksiselitteisesti, että saatu vaikutus on kokeessa riippumattomaksi muuttujaksi määritellyn tekijän tuottama. Tietyn toimenpiteen vaikuttavuudella tarkoitetaan siis sitä, että mikään muu seikka ei ole tuottanut havaittua lopputulosta eli vaikutusta. Hoitoyhteisissä tämä merkitsee, että sattuman, spontaanin paranemisen, lumevaikutuksen tai jonkin rinnakkaihoidon sekoittavan vaikutuksen mahdollisuus on voitu sulkea pois. Vaikuttavuustutkimuksessa sekä riippumattoman muuttujan että hoitotuloksen määrittely edellyttää yleensä, että on olemassa perusteltu käsitteellinen malli hoidon vaikutusmekanismeista.

Vaikuttavuustutkimusta, kuten muutakin tieteellistä tutkimusta, voidaan tehdä monenlaisilla tutkimusasetelmilla [1]. Satunnaistettu vertailukoe (randomised controlled trial, RCT) on keskeinen menetelmä tarkastella vaikuttavuutta, sairauksien ehkäisyä, hoitoa, kuntoutusta tai vaikkapa uuden lääkeaineen tehoa ja turvallisuutta. Menetelmää pidetään luotettavimpana tapana mitata vaikuttavuutta [2].

Satunnaistettu vertailukoe voidaan määritellä monin tavoin, mutta tavallisesti se kuvataan tutkimukseksi, jossa joukko tutkit-

tavia jaetaan satunnaisesti kahteen tai useampaan ryhmään. Yksi näistä ryhmistä saa tutkittavaa interventiota ja muut muuta hoitoa (kuten olemassa olevaa standardihoitoa), plaseboa tai ei mitään.

Käyttämällä satunnaistettua vertailukoetta pyritään välttämään ongelmat, jotka liittyvät ei-kontrolloituihin tai ”historiallisia kontrolloja” käytäviin kokeisiin. Kontrolloimattomissa kokeissa voidaan saada hyvinkin ylioptimistisia tuloksia intervention vaikutuksesta, koska vaikutusta ei verrata minkään toisen ryhmän tilanteeseen. Historiallisten kontrollien käytöllä tarkoitetaan tilannetta, jossa tutkija haluaa välttää satunnaistuksen poimimalla takautuvasti esimerkiksi potilasrekisteristä aiemmin standardihoitoa saaneita potilaita ja käyttää tätä ryhmää uuden intervention verrokkiryhmänä. Tähän liittyy monia ongelmia, joista tärkein on tarkasteltavien ryhmien heikko vertailukelpoisuus.

Seuraavaksi tarkastelen yleisiä tilastollisia asioita, joita satunnaistetun vertailukokeen suunnittelussa tulisi huomioida [3,4]. En perehdy tilastollisiin menetelmiin tai mallinnoihin enkä ulota tarkastelua integroidun tutkimusdatan eli metadatan erityispiirteiden vaikutusten huomioimiseen.

## KOHDEJOUKKO

Kohdejoukko tarkoittaa sitä ryhmää, jota tutkimuksen tuloksista tehdyt johtopäätökset yleistetään koskemaan. Tutkimukseen rekrytoitavien tutkimushenkilöiden tulisi edustaa mahdollisimman hyvin kohdejoukkoa, ainakin varmentavissa eli konfirmatorisissa tutkimuksissa. Jos tutkimushenkilöt valitaan hyvin tiukoin sisäänottokriteerein, he edustavat vain pientä osaa siitä kohdejoukosta, jolle interventio halutaan suunnata. Tästä syystä esimerkiksi myyntilupaa uudelle lääkkeelle ei voi saada, jos sitä on tutkittu vain nuorilla miehillä, joilla ei ole yleisimpiä kroonisia sairauksia.

Myös maantieteellisissä rajauksissa on syytä olla tarkkana. Jos esimerkiksi uusi nuorten aikuisten etähoito-ohjelma on todettu vaikuttavaksi harvaanasutuilla seuduilla, se ei ole välttämättä vaikuttava koko Suomessa.

Kohdejoukkoon liittyvää otoskokoa ja siihen vaikuttavia tekijöitä käsitellään tarkemmin edempänä, mutta yleisesti voidaan sanoa, että mitä suurempi kohdejoukko, sitä suurempi määrä tutkimushenkilöitä tutkimukseen tarvitaan. Jos kohdejoukko on hyvin laaja, voi olla hankalaa tilastollisin menetelmin arvioida tarvittavaa rekrytoitavien määrää. Lisäksi olemassa olevan intervention (vertailu-interventio) vaikutuksen suuruudesta ei välttämättä löydy julkais-tua tietoa. Toisaalta liialliset poissulku- ja sisäänottokriteerit voivat vaikeuttaa rekrytointiprosessia, jolloin tutkimukseen tarvittavaa otoskokoa voi olla lähes mahdotonta saavuttaa sovitussa ajassa ja käytettävissä olevilla resursseilla.

## VERROKKIRYHMÄT

Tutkittavien poissulku- ja sisäänottokriteerien määrittämisen jäl-keen tutkijan on valittava vertailtavat ryhmät eli tarkasteltavat inter-ventiot. Useimmiten satunnaistetussa vertailukokeessa verrataan tietyn intervention, kuten uuden terapiamuodon, vaikutuksia ole-massa olevan hoitomuodon vaikutuksiin. Tällöin tärkeää on tarkas-tella, löytyykö olemassa olevasta hoidosta tutkittua tietoa vai ei. Jos tietoa ei ole, tutkija ei voi olla varma, kuinka tehokas tai vaikuttava uusi interventio todella on, vaikka tutkimuksen tulokset tukisivat

## ”Kontrolloimattomissa kokeissa voidaan saada ylioptimistisia tuloksia intervention vaikutuksesta.”

väitettä, että uusi interventio eroaa merkitsevästi olemassa olevasta hoidosta. Saattaahan olla, että olemassa oleva hoito ei merkittävästi eroa tilanteesta, jossa tutkittavat eivät saa mitään hoitomuotoa.

Engelman voi ratkaista esimerkiksi ottamalla tutkimukseen kol-me rinnakkaisryhmää, joista yksi saa uutta tutkittavaa interventiota, toinen standardihoitoa ja kolmas ei saa mitään hoitoa (ns. ”puhdas” verrokki).

## VASTEET

Ensisijaisen vasteen eli päämuuttujan (primary variable) pitäisi olla välittömässä yhteydessä tutkimuksen päätarkoitukseen. Valinnan tulisi heijastaa tutkimusalan hyväksytyjä normeja ja standardeja. Suositus on, että valittava muuttuja on luotettava ja validoitu aiempien tutkimusten tai kirjallisuuden perusteella. Myös siitä pitäisi olla näyttöä, että päämuuttujan avulla saadaan merkityksellistä ja merkittävää tietoa tutkittavan (potilas)ryhmän hoitamisesta. Esimerkiksi tutkittaessa ahdistushäiriön uuden, nuorille suunnatun hoitomuodon vaikuttavuutta päämuuttujan tulisi olla kehitetty ja yleisesti käytetty mittaamaan juuri ahdistusta eikä esimerkiksi masennusta tai syömishäiriötä.

Tutkimuksen otoskoko lasketaan yleensä ensisijaiselle vasteelle. Tavallisesti satunnaistetussa vertailukokeessa valitaan vain yksi tehoa, turvallisuutta, elämänlaatua tai kustannusvaikuttavuutta mittaava vaste. Poikkeuksena ovat ns. monimutkaiset interventiot, joissa joudutaan käyttämään useita ensisijaisia vasteita [ks. 5,6]. Mi-käli ensisijaisia vasteita on enemmän kuin yksi, on mietittävä tyyppin I virheen vaikutuksen kontrollointia, kuten myös vaikutusta tyyppin II virhetasoon ja otoskoko.

Tyyppin I virhe syntyy, kun tutkimusaineisto todellisuudessa tukee nollahypoteesia (H0) tarkasteltavan vasteen suhteen, mutta käytetty tilastollinen testi hylkää sen (taulukko 1). Toisin sanoen tarkasteltavat interventiot eivät todellisuudessa eroa toisistaan tarkasteltavan

## TAULUKKO 1.

### Tyyppin I ja II virheet

Todellinen tilanne	Johtopäätös		
	H0	Oikea johtopäätös	HV
H0	Oikea johtopäätös	Tyyppin I virhe ( $\alpha$ )	
HV	Tyyppin II virhe ( $\beta$ )	Oikea johtopäätös	

H0 = nollahypoteesi

HV = vaihtoehtoinen hypoteesi

vasteen suhteen, mutta tilastollinen testi antaa merkitsevyydystason alittavan p-arvon, jolloin vaihtoehtoinen hypoteesi (HV) hyväksytään. Yleensä tyyppin I virheelle käytetään riskitasoa  $\alpha = 0,05$ .

Tyyppin II virhe syntyy, kun tutkittavat hoitomallit todellisuudessa eroavat toisistaan, mutta tilastollisen testin tulos ei tätä tue. Esimerkiksi kahden ensisijaisen vasteen tutkimuksessa monivertailuongelma voidaan ratkaista käyttämällä ns. Bonferroni-merkitsevyydystasokorjausta. Käytännössä tavoiteltu merkitsevyydystaso jaetaan testien lukumäärällä, ja näin laskettua arvoa käytetään jokaisessa yksittäisessä testissä. Esimerkiksi kahden muuttujan tapauksessa arvon  $\alpha = 0,05$  sijasta käytetäänkin arvoa  $\alpha = 0,025$ . On kuitenkin tärkeää huomioida, että analyysien voima heikkenee monivertailun huomioimisen myötä. Toisin sanoen todennäköisyys tyyppin II virheelle kasvaa ja oikeita eroavaisuuksia voi jäädä havaitsematta.

Toissijaiset vasteet (secondary variables) joko tukevat tutkimuksen päätarkoitusta tai ovat yhteydessä tutkimuksen toissijaisiin tavoitteisiin. Näidenkin vasteiden määrittely, arvioinnin ja analysoinnin pitäisi olla etukäteen kirjattuna tutkimussuunnitelmaan. Toissijaiset vasteet voivat olla paitsi yksittäisiä muuttujia myös ns. yhdistelmämuuttujia (composite variables), eli useat yksittäiset kysymykset voidaan yhdistää yhdeksi muuttujaksi protokollassa etukäteen määritetyn algoritmin mukaisesti.

Menetelmiä yhdistelmämuuttujien muodostamiseksi on useita. Tavallisesti ne toteutetaan laskemalla muuttujien arvoista aritmeettinen tai painotettu keskiarvo tai ryhmittämällä muuttujat merkityksensä perusteella muutamaaan isompaan ryhmään. Lisäksi on mahdollista käyttää vasteena yleisarviota (global assessment), joka sisältää useimmiten tutkijan subjektiivisen arvon. Tällöin on erityisen tärkeää, että suunnitelmaan on etukäteen kirjattu vastausvaihtoehdojen perusteet. Jos suoran vaikutuksen arviointi ei ole mahdollista, voidaan käyttää sijaisvasteita (surrogate variables).

Jatkuvien muuttujien luokittelu kategorioihin tavallisesti johtaa uusien muuttujien informaation vähenemiseen, mikä heikentää tilastollisten testien voimaa. Tämä pitäisi huomioida otoskoon arvioinnissa. Jatkuvia muuttujia voivat olla esimerkiksi käytetty rahamäärä tai tutkimushenkilöiden digitaaliseen mielenterveyttä edistävään ohjelmaan käyttämä aika.

## KOEASETELMAT JA SATUNNAISTAMINEN

Yleisimmin satunnaistettu vertailukoe toteutetaan rinnakkaisten ryhmien koeasetelmana. Asetelmassa on tavallisesti kaksi ryhmää ja tutkimushenkilöt satunnaistetaan näistä jompaankumpaan. Ristikkäiskaavio- tai crossover-koeasetelmassa tutkittavat satunnaistetaan käsittelyjonoihin. Asetelma sopii parhaiten esimerkiksi jonkin uuden hoitomuodon vaikuttavuuden vertaamiseen standardihoitoon kroonisesti sairailta henkilöillä. Ideana on, että tutkittava saa molemmat hoitomuodot satunnaisessa järjestyksessä. Tutkittava

toimii siis itse itsensä verrokkina. Oleellista on, että puhdistautumisjakso (washout period) hoitomuotojen välillä on riittävän pitkä, jotta ensimmäinen hoitoperiodi ei vaikuttaisi ns. jäännösvaikutuksena toisen hoitoperiodin tuloksiin.

Kolmantena koeasetelmatyypinä ovat faktorikokeet, joista yksinkertaisin on 2 x 2 -faktorimalli. Siinä tutkittavia käsitteilyt on kaksi (merkitään A ja B) ja tutkittavat satunnaistetaan neljästä ryhmästä yhteen. Kahdesta käsitteilystä muodostetaan siis neljä vaihtoehtoa: 1) pelkästään käsittely A, 2) pelkästään käsittely B, 3) käsittely A ja B yhdessä sekä 4) ei mitään.

2 x 2 -faktorikokeen etuna on se, että samalla tutkittavien lukumäärällä saadaan joissakin tapauksissa arvioitua yhden interventio sijasta kaksi interventiota. Eli vaikka tutkitaan kahta interventiota niin otoskooksi riittää sama kuin olisi tarvittu tutkittaessa vain yhtä interventiota. Oletuksena on, että tutkittavat interventiot ovat riippumattomia eli niillä ei ole merkitsevää yhdysvaikutusta. Toisin sanoen käsittelyiden A ja B vaikutus päävasteeseen on luonteeltaan additiivista.

Additiivisessa vaikutuksessa interventioiden yhteisvaikutus on siis likimain yhtä suuri kuin osavaikutusten summa. Jos otoskoko lasketaan interventioiden A ja B vaikutuksille, yhdysvaikutuksen testistä voi tulla hyvin heikko. Näin ollen tutkimukselle on tärkeää määrittää riittävä otoskoko, jotta mahdollinen yhdysvaikutus voidaan havaita luotettavasti tilastollisin testein tutkimusaineistosta.

Kaikissa tutkimusasetelmissa yksilösatunnaistaminen ja -kontrollointi eivät ole mahdollisia, vaan joudutaan etsimään muita ratkaisuja. Suunnittelun kannalta haasteellisia tutkimusasetelmia käytetään mm. monimuotoisissa ja laajoissa palvelujärjestelmä- ja interventiotutkimuksissa. Erityisen haasteen muodostaa toimintaympäristön vaikutus, jota on tutkimusta suunniteltaessa vaikea ennakoita.

Vaihtoehtona yksilösatunnaistamiselle on muodostaa ryppäitä ja satunnaistaa nämä ryppäät tutkimusryhmiin (cluster randomisation). Käytännössä otos kerätään kaksivaiheisesti eli poimitaan ensin joukko ryppäitä kaikkien ryppäiden joukosta. Seuraavassa vaiheessa poimitaan jokaisesta ryppästä osaotos ja yhdistetään osaotokset yhdeksi otokseksi.

Tällaista ryväsotantaa (kuvio 1) käytetään esimerkiksi kouluissa tehtävissä interventiotutkimuksissa, joissa tutkitaan vaikkapa

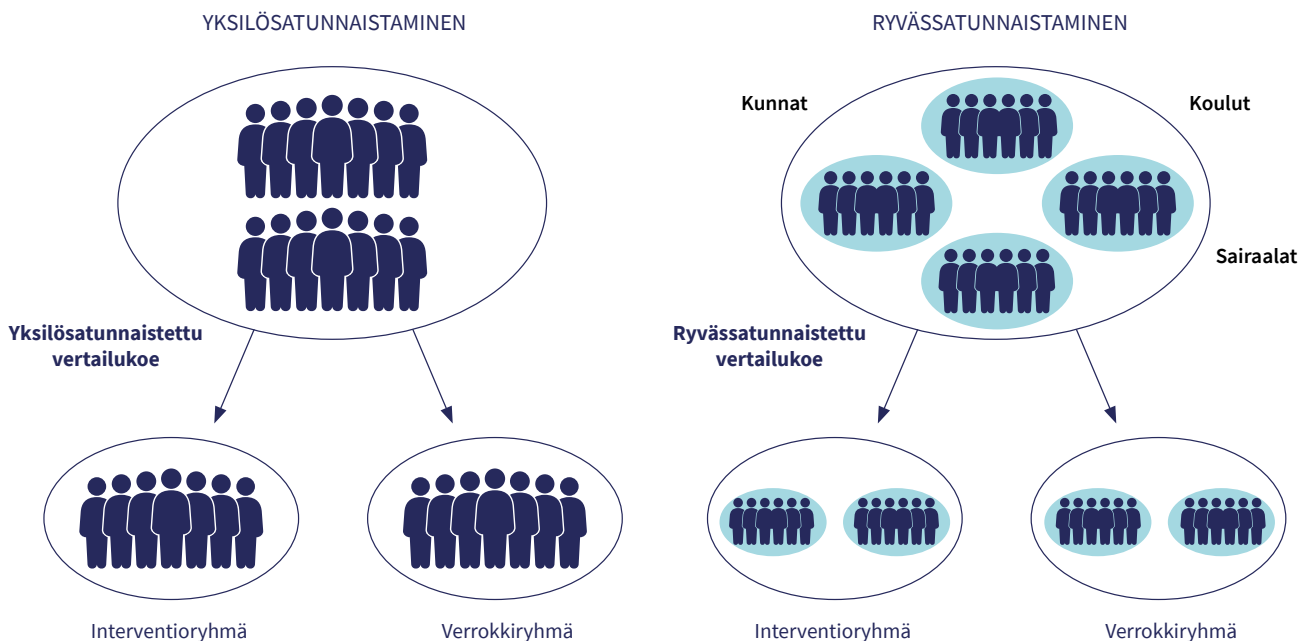
**”Kohdejoukko tarkoittaa sitä ryhmää, jota tutkimuksen tuloksista tehdyt johtopäätökset yleistetään koskemaan.”**

uuden mielenterveyteen liittyvän ohjelman vaikutuksia kouluki- saamiseen. Moniasteisessa ryväsotannassa perusjoukko voidaan jakaa ryppäisiin hierarkkisesti eli perusjoukko jaetaan ensin ryppäisiin, jotka puolestaan voidaan jakaa aliryppäisiin jne. Esimerkki tällaisesta asetelmasta voisi olla mielenterveysohjelmia vertaileva tutkimus, johon osallistuisi yksittäisiä koululuokkia eri kouluista. Hierarkkinen rakenne voisi tällöin olla maakunnat, kunnat, koulu- piirit, koulut ja luokat. Tällaiset tutkimukset vaativat suuremman määrän tutkittavia ja ovat huomattavasti monimutkaisempia toteuttaa sekä analysoida kuin yksilösatunnaistamisen kautta tehtävät tutkimukset.

Tutkimushenkilöiden valikoitumista ja siihen liittyvää harhaa pyritään välttämään satunnaistamisen (randomisation) lisäksi sokkouttamisella (blinding). Korkein sokkouttamisen taso, jossa

KUVIO 1.

### Ryväsotanta



kukaan tutkimuksessa osallisena oleva, mukaan lukien tutkittava itse, ei tiedä mitä interventiota yksittäinen tutkimushenkilö saa, on useimmiten mahdollon toteuttaa. Realistisempaa on, että tutkimuksen statistikko tai tilastollisista analyyseistä vastaava henkilö analysoi esimerkiksi ensisijaisen vasteen sokkona, jonka jälkeen muu aineisto voidaan tutkimusryhmien osalta avata.

Satunnaistamisella pyritään siis siihen, että ryhmät ovat samankaltaisia muiden kuin tutkittavien vaikuttavien tekijöiden osalta. Tavallisimmin satunnaistaminen toteutetaan arpomalla kuhunkin

## *”Päämuuttujan pitäisi olla välittömässä yhteydessä tutkimuksen päätarkoitukseen.”*

ryhmään yhtä monta tutkimushenkilöä (1:1 allokointi). Joissakin tapauksissa kannattaa harkita kaksi kertaa suurempaa tutkimushenkilöiden määrää uutta interventiota saavien ryhmään kuin verrokki- tai esimerkiksi käytössä olevaa interventiota saavien ryhmään (2:1 allokointi). Näin saadaan kerättyä enemmän tietoa tutkitavan intervention vaikuttavuudesta ja turvallisuudesta.

Tutkijat harvoin perustelevat julkaisuissaan tätä epätasaista satunnaistusta tai kertovat, miten epätasainen satunnaistus on huomioitu otoskoon laskemisessa. Perusteita kannattaa kuitenkin miettiä: onko kyse esimerkiksi 1) kustannusten minimoimisesta, 2) harvinaisempien haittatapahtumien löytämisestä isommalla otoskolla, 3) toiveesta saada rekrytointi tehtyä nopeammin tai 4) arviosta uuden intervention ryhmän suuremmasta keskeyttämisasteesta.

Jotta 2:1 allokoinnilla saataisiin sama tilastollinen voima kuin 1:1 allokoinnilla, tutkimushenkilöitä pitää rekrytoida keskimäärin yli 10 % enemmän [7]. Vastaavasti 3:1 allokoinnilla tarvitaan keskimäärin yli 30 % enemmän tutkimushenkilöitä. Tutkittavan hoitomuodon, tutkimuskohteen ja kohdejoukon huomioiden tutkimuksen satunnaistaminen 2:1 voi mielestäni joskus olla eettisesti arveluttavaa. Tällöin perustelluinta on pitäytyä 1:1 allokoinnissa.

Toisinaan eettisesti hyvänä vaihtoehtona pidetään satunnaistamista tutkitavaan interventioon tai saman intervention odotuslistalle. Odotuslistaryhmä on siis intervention verrokkiryhmä (wait list control group) [8]. Odotuslistalle satunnaistetut saavat tutkitavan intervention tutkimussuunnitelmassa määritellyn ajan kuluttua. Asetelmassa intervention vaikutusta arvioidaan vertaamalla muutosta interventioryhmässä alkutilanteesta intervention loppuun muutokseen odotuslistaryhmässä alkutilanteesta hetkeen ennen intervention aloittamista.

Tutkimusten mukaan odotuslistaryhmän käyttäminen voi keinotekoisesti lisätä intervention vaikutuksen suuruutta, sillä hoidon odottaminen passivoi. Jos ihminen ei olisi odotuslistalla, hän saattaisi aktiivisesti etsiä parannusta. Koeasetelmaa käytetään yhä enemmän psykologisissa ja käyttäytymistieteiden interventio-tutkimuksissa. Tutkijan kannattaa tarkoin miettiä koeasetelman sopivuutta omaan tutkimusongelmaansa ja kohdejoukkoonsa. Asetelma myös väistämättä pidentää tutkimuksen kestoa. Tilastollinen analyysi ei sinänsä eroa tavallisesta kahden rinnakkaisryhmän analyysistä.

Satunnaistaminen on hyvä tehdä jonkin tarkoitukseen soveltuvan tilastollisen ohjelman avulla. Näin taataan, että satunnaistus on tarvittaessa toistettavissa. Mikäli tutkimuksessa on useita eri alueita kuten sairaanhoitopiirejä, kuntia tai kouluja, kannattaa harkita ositettua otantaa, jossa kullekin osajoukolle tehdään omat satun-

naistamislistansa. Käytännössä tämä tarkoittaa, että jos tutkimusta tehdään esimerkiksi neljässä sairaanhoitopiirissä ja halutaan osittaa otanta sairaanhoitopiiriin mukaan, kullekin sairaanhoitopiirille muodostetaan oma satunnaistamislistansa, eli yhteensä tehdään neljä listaa. Näin taataan, että tarkasteltavat ryhmät ovat likimain yhtä suuria kullakin alueella.

Tutkimus on luonteeltaan joko konfirmatorinen tai eksploratiivinen. Ensimmäisessä tutkimushypoteesit on päätetty etukäteen, ja niillä on suora yhteys tutkimuksen päätarkoituksiin. Tilastollisin testeillä saadaan analysoitua tilastolliset merkitsevyydet, jotka tavallisesti raportoidaan yhdessä vaikutuksen suuruuden (effect size) kanssa. Näiden lisäksi tutkimuksen johtopäätöksiin vaikuttaa aina myös kliininen/käytännön merkitsevyys.

Konfirmatoristen tutkimusten tarkoituksena on tuottaa selkeitä todisteita tutkitavan hoidon tehosta tai turvallisuudesta. Konfirmatorisia tutkimuksia edeltää tavallisesti joukko eksploratiivisia tutkimuksia, joissa niin ikään on selkeästi määritellyt tavoitteet, muttei etukäteen määriteltyjä hypoteeseja. Näin ollen tilastollinen analyysikin on lähinnä kuvailevaa.

Yksi esimerkki eksploratiivisesta tutkimuksesta on toteutettu vuustutkimus (feasibility study), jossa tarkoituksena voi olla vaikka tutkia uuden digitaalisen mielenterveyden lukutaito-ohjelman sopivuutta ammattiin opiskeleville 16–18-vuotiaille. Eksploratiivisissa tutkimuksissa saatujen tulosten luotettavuus on hyvä aina todentaa konfirmatorisin tutkimuksin. Huomioitavaa on, että konfirmatorisissa tutkimuksissa voi hyvin olla mukana luonteeltaan eksploratiivisia vasteita.

### **VERTAILUTYYPIT**

Hoitomuodon toimivuus, vaikuttavuus tai teho on tieteellisesti vakuuttavimmin osoitettavissa superioriteettitutkimuksissa. Niissä tarkoituksena on osoittaa kahden eri hoitomuodon (tai uutta hoitoa saavan ryhmän ja verrokkiryhmän) eroavaisuus toisistaan. Vertailtava hoitomuoto (active comparator) olisi hyvä valita huolella. Esimerkkinä sopivasta vertailuinterventiosta on samaan indikaatioon laajasti käytetty terapia, jonka vaikuttavuus on osoitettu hyvin suunnitelluissa ja raportoiduissa superioriteettitutkimuksissa. Mahdollista on myös osoittaa (kliininen) ekvivalenssi tai non-inferioriteetti kahden hoitomuodon välillä. Kuvio 2 havainnollistaa

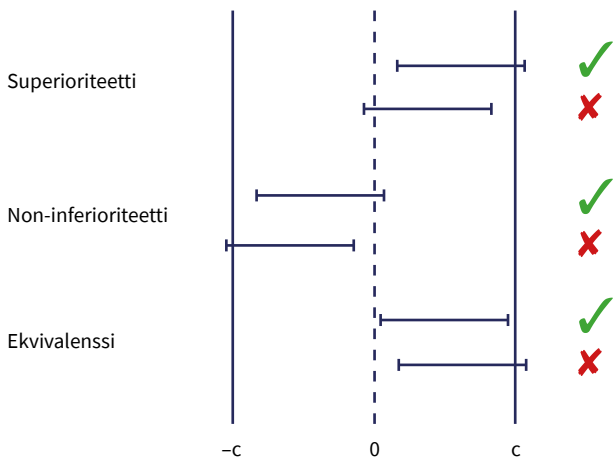
## *”Rekisteröinnin tavoitteena on saada kaikki tutkimukset tuloksineen julkisiksi.”*

näitä kolmea vertailutyyppiä. Tutkijan kannalta oleellista on erottaa, miten vertailutypit eroavat toisistaan – mitä hypoteesia missäkin niistä testataan – ei niinkään miten estimaatit luottamusväleineen eri tyyppisille vasteille teknisesti lasketaan kahden hoitomuodon eron välille.

### **OTOSKOKO**

Tutkittavien lukumäärän pitäisi aina olla riittävän suuri, jotta tutkimuksella pystyttäisiin vastaamaan suunnitelmassa esitettyihin tutkimusongelmiin. Otokoko perustuu tavallisesti tutkimuksen ensisijaiseen tarkoitukseen ja se lasketaan käyttäen kirjallisuudesta saatuja tietoja ensisijaisesta vasteesta. Otokoko voidaan laskea, kun tutkimussuunnitelmassa on kirjattuna ensisijainen vaste, tilastollinen testi ja nollahypoteesi sekä sen vaihtoehdot hypoteesit.

### Superioriteetti, non-inferioriteetti ja ekvivalenssi -vertailutyypit



Kasvun tuki  
© 2021

Jokaisesta asetelmasta on kuvattu kaksi esimerkkiä 95 %:n luottamusvälistä kahden käsittelyn vaikutuksen erolle. Vihreä merkintä tarkoittaa, että nollahypoteesi ei saa tukea ja punainen että se saa tukea Non-inferioriteetin päättelmissä arvo  $-c$  on käytettävä nollahypoteesin arvo, ja ekvivalenssin laskemisessa arvot  $-c$  ja  $c$  ovat käytettävät nollahypoteesin arvot.

**Superioriteetti:** Ensimmäisessä (vihreä) esimerkissä nollahypoteesi "käsittelyiden välillä ei eroa vaikuttavuudessa" hylätään eli data ei tue nollahypoteesia. Luottamusväli kahden käsittelyn erolle ei sisällä arvoa 0. Toisessa esimerkissä (punainen) käsittelyiden eron luottamusväli sisältää arvon 0, joten nollahypoteesi "käsittelyiden välillä ei eroa vaikuttavuudessa" saa tukea.

**Non-inferioriteetti:** Tarkoitus on osoittaa, että uusi hoitomuoto ei ole vallitsevaa hoitomuotoa huonompi. Nollahypoteesina on, että "käsittelyn A vaikutus on huonompi kuin käsittelyn B". Ensimmäisessä esimerkissä (vihreä) käsittelyiden eron luottamusvälin alaraja on suurempi kuin  $-c$ , joten nollahypoteesi hylätään eli käsittely A ei ole huonompi kuin käsittely B. Toisessa esimerkissä (punainen) käsittelyiden eron luottamusvälin alaraja on pienempi kuin  $-c$ , joten nollahypoteesi saa tukea eli käsittelyn A vaikutus näyttää olevan käsittelyn B vaikutusta huonompi. Non-inferioriteetin vallitessa luottamusvälin alaraja ei siis saa olla sallittua eroa pienempi. Lisäksi on huomioitava, että arvo  $c$  päätetään tutkimuksen suunnitteluvaiheessa ja kirjataan suunnitelmaan. Olisi virheellistä päätellä tilastollisesta testistä saadun ei-merkitsevän  $p$ -arvon perusteella, että non-inferioriteetti toteutuu kahden intervention välillä.

**Ekvivalenssi:** Ensimmäisessä (vihreä) esimerkissä nollahypoteesi "käsittelyiden vaikutukset eivät ole ekvivalentit" ei saa tukea, koska luottamusväli kahden käsittelyn erolle sisältyy välille  $[-c, c]$ . Toisessa (punainen) esimerkissä käsittelyiden eron luottamusvälin yläraja on suurempi kuin  $c$ , joten ekvivalenssi ei toteudu (nollahypoteesi saa tukea). Tähän pätee sama sääntö kuin non-inferioriteetinkin tapauksessa eli arvot  $-c$  ja  $c$  päätetään ennen tutkimuksen aloittamista ja kirjataan tutkimussuunnitelmaan. Myös tässä tapauksessa olisi virheellistä päätellä tilastollisesta testistä saadun ei-merkitsevän  $p$ -arvon perusteella, että ekvivalenssi toteutuu kahden intervention välillä.

Tutkimuksen tulokset voidaan tulkita hyvin eri tavoin riippuen siitä, onko kyse superioriteetin, non-inferioriteetin vai ekvivalenssin osoittamisesta. Tietyin ehdoin on mahdollista vaihtaa superioriteetin osoittamisesta non-inferioriteetin osoittamiseen tai päinvastoin, vaikka tutkimus olisi jo käynnissä [9].

Suunnitelmaan kirjoitetaan otoskoon määrittämisen kohdalle 1) mikä vaste, 2) mitä arvoja on käytetty (keskiarvo, hajonta, vastausprosentti, eron estimaatti), 3) mitä otoskoon laskemismenetelmää on käytetty, 4) miten tutkittavien keskeyttämiset, vetäytymiset ja protokollasta poikkeamiset on huomioitu ja 5) tyyppin I ja II virhetasot. Tavallisimmin tyyppin I virhe kiinnitetään tasolle  $\beta = 0,05$  ja tyyppin II virhe välille  $0,1 \leq \beta \leq 0,2$ . Tilastollisen testin voima määritetään kaavalla  $(1 - \beta) \times 100\%$  eli se vaihtelee välillä 80–90 %.

### TUTKIMUKSEN REKISTERÖINTI

ICMJE:n (International Committee of Medical Journal Editors) kriteerien mukaan rekisteröitäviksi interventiotutkimuksiksi katsotaan prospektiiviset (etenevät) ja kontrolloidut (vertailevat, joissa käytössä on ainakin kaksi eri hoitoa; koe- ja vertailumenetelmä) lääkkeillä, laitteilla, kirurgisilla tai käyttäytymis- ja muilla hoitomenetelmillä tehtävät hoitotutkimukset, joissa arvioidaan käytettyjen hoitomenetelmien vaikuttavuutta [10].

Edellä mainitut ehdot täyttävät satunnaistetut vertailukokeet on aina rekisteröitävä. Käytetyin julkinen, ilmainen ja kaikille avoin, ei-kaupallinen tutkimusrekisteri on tällä hetkellä ClinicalTrials.gov (<https://clinicaltrials.gov>). ICMJE:n periaatteiden mukaan lehdet ottavat julkaistavaksi vain rekisteröityjä tutkimuksia. Rekisteröinnin tavoitteena on saada kaikki tutkimukset tuloksineen julkisiksi, lisätä yleistä luottamusta tutkimustoimintaan ja välttää päällekkäisiä tutkimuksia eettisyyden nimissä. Tutkimuksen rahoituslähde ei vaikuta rekisteröintivelvoitteen eli myös ns. tutkijalähtöiset tutkimukset on rekisteröitävä ennen ensimmäisen tutkimushenkilön ottamista mukaan tutkimukseen.

### KAIKKIVOIPA SATUNNAISTETTU VERTAILUKOE?

Nykykaikaisen satunnaistetun kontrolloidun tutkimusasetelman "isänä" pidetään Austin Bradford Hilliä, joka julkaisi ryhmänsä kanssa vuonna 1948 ensimmäisen satunnaistetun vertailukokeensa tulokset. Sittemmin asetelman käyttö on levinnyt eri tieteenaloille. Suurin vahvuus on tutkittavien erilaisesta valikoitumisesta tutkimusryhmiin aiheutuvan harhan hallitseminen.

Asetelma sopii jokseenkin kaikille tieteenaloille. Joidenkin mielestä poikkeuksena ovat tutkimukset, joissa ei ole eettisesti hyväksyttävää riistää potilailta vaikuttaviksi osoitettuja hoitomuotoja. Mutta miten tämä määritellään? Mitkä ehdot yksittäisen tutkimuksen pitäisi täyttää, jotta satunnaistettu vertailukoe ei olisi toteutettavissa? Tietääkseni tästä ei ole yleisesti hyväksyttyä "kultaista" sääntöä, vaan ainoastaan joukko mielipiteitä.

Silti voidaan sanoa, että satunnaistettua vertailukoetta ei pidä tehdä, jos johtopäätös tiedetään suurella todennäköisyydellä etukäteen. Silloinhan se on resurssien tuhlaamista, eikä tutkittavien edun mukaista. Tämänhetkinen ehdotus kriteereistä sisältää seuraavat kolme kohtaa: Jokaisessa satunnaistetussa vertailukokeessa tulisi olla 1) selkeä kysymys 2) yksityiskohtaisesti muodostettu tutkimushypoteesi ja näiden lisäksi 3) olisi varmistettava, että olemassa olevan tiedon mukaan hypoteesia ei ole vielä tutkittu (ei myöskään ei-satunnaistetuilla kokeilla).

Satunnaistetun vertailukokeen suunnittelussa on monia asioita, jotka tutkijan on huomioitava. Suunnittelu vaatii moniosaamista, joten tutkimusryhmällä on syytä olla riittävästi tietoa myös tutkimuslainsäädännöstä, tutkimuslupaprosessista, eettisistä asioista, tilastoiikasta ja tutkimusdatan keräämisestä. Toisaalta täydellistä satunnaistettua vertailukoetta ei uskoakseni ole koskaan suunniteltu ja toteutettu. Riittää, että suunnittelun "palaset" ovat riittävän hyvin paikoillaan. ●

## AVAINSANAT:

---

vaikuttavuus, RCT, satunnaistettu vertailukoe, satunnaistaminen, interventio

## SIDONNAISUUDET:

---

Sanna Hinkka-Yli-Salomäki: Ei sidonnaisuuksia.

## LÄHTEET:

---

- 1 Guyatt GH, Oxman AD, Kunz R ym. ja GRADE Working Group (2008). What is "quality of evidence" and why is it important to clinicians?. *BMJ* 336, 995–8. <https://doi.org/10.1136/bmj.39490.551019.BE>
- 2 Hariton E & Locascio JL (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG* 125, 1716. <https://doi.org/10.1111/1471-0528.15199>
- 3 Guidance for clinical trial protocols. SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials). <https://www.spirit-statement.org/>
- 4 European Medicines Agency. ICH Topic E9: Statistical principles for clinical trials. ICH Harmonised Tripartite Guideline. [www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials](http://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials)
- 5 Medical Research Council (2021). Developing and evaluating complex interventions. [www.mrc.ac.uk/complexinterventionsguidance](http://www.mrc.ac.uk/complexinterventionsguidance)
- 6 Craig P, Dieppe P, Macintyre S ym. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 337, a1655. <https://doi.org/10.1136/bmj.a1655>
- 7 Spencer Phillips Hey & Jonathan Kimmelman (2014). The questionable use of unequal allocation in confirmatory trials. *Neurology* 82(1): 77-79. doi: 10.1212/01.wnl.0000438226.10353.1c
- 8 Cunningham JA, Kypri K, McCambridge J (2013). Exploratory randomized controlled trial evaluating the impact of a waiting list control design. *BMC Med Res Methodol* 13, 150. <https://doi.org/10.1186/1471-2288-13-150>
- 9 Committee for Proprietary Medicinal Products (2001). Points to consider on switching between superiority and non-inferiority. *Br J Clin Pharmacol* 52, 223–228. <https://doi.org/10.1046/j.0306-5251.2001.01397-3.x>
- 10 ICMJE. Clinical Trials Registration. [www.icmje.org/about-icmje/faqs/clinical-trials-registration/](http://www.icmje.org/about-icmje/faqs/clinical-trials-registration/)