

# Satunnaistetun vertailukokeen laadukkaan raportoinnin edellytykset

Artikkeli on jatkoa Kasvun tuki -aikakauslehdessä 2/2021 julkaistulle artikkelille Vaikuttavuustutkimus: Satunnaistettu vertailukoe tulee suunnitella huolella.

- Satunnaistetun vertailukokeen toteuttaminen on suuri ponnistus, joka vaatii käytännön asioiden järjestelmistä ja perehtymistä tutkimuksen laadukkaaseen raportointiin.
- Tutkimusartikkelin kirjoittamisessa auttaa CONSORT 2010 -tarkistuslistan noudattaminen.
- Yli puolessa raportoiduista satunnaisista vertailukokeista on käytetty kahden rinnakkaisryhmän koeasetelmaa.
- Ennen vertailukokeen toteuttamista kannattaa tehdä pilottitutkimus.
- Tulokset esitetään ITT-periaatteen (intention-to-treat) mukaisesti, eli kaikki satunnaistetut tutkimushenkilöt pidetään mukana.



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

Satunnaistettu vertailukoe määritellään tavallisesti tutkimukseksi, jossa joukko tutkittavia jaetaan satunnaisesti kahteen tai useampaan ryhmään. Yksi näistä ryhmistä saa tutkittavaa interventiota ja muut ryhmät muuta hoitoa, kuten olemassa olevaa tavallista hoitoa, tai eivät mitään. Kun halutaan tutkia intervention vaikutusta (esimerkiksi tehoa, sopivuutta tai paremmuutta yleensä) tietyssä kohdejoukossa, on tärkeää verrata interventiota saavia heihin, jotka saavat muuta hoitoa tai eivät saa mitään. Satunnaistetun vertailukokeen suurin vahvuus on tutkittavien tutkimusryhmiin valikoitumisesta aiheutuvan harhan hallitseminen (Hinkka-Yli-Salomäki, 2021). Siksi tämä koeasetelma on jo pitkään ollut vaikuttavuustutkimusten ns. kultainen standardi.

## PILOTOINTIVAIHE

Satunnaistetun vertailukokeen tekeminen on tutkimusryhmälle suuri ponnistus. Tutkimussuunnitelman huolellisen kirjoittamisen lisäksi tarvitaan monien käytännön asioiden järjestelmistä sekä riittävää perehtymistä tutkimuksen raportointiin.

Ennen kokeen läpiviemistä se kannattaa pilotoida. Pilotointivaiheessa voidaan esimerkiksi:

1. Tarkastella vertailukokeeseen suunniteltavien mittareiden ja kysymysten sopivuutta sekä mahdollisten toimenpiteiden toteutettavuutta ja hyväksyttävyyttä:
  - Onko kysely ajallisesti sopivan mittainen?
  - Saadaanko kysymyksillä tarvittava tieto tutkimuksen tavoitteisiin vastaamiseksi?
  - Onko kysymyksiä liikaa?
  - Ovatko kysymysajankohdat oikein valitut ja onko ajankohtien aikaikkuna käytännössä toimiva?
  - Onko suunnitellut toimenpiteet mahdollista toteuttaa niille varatussa ajassa?
2. Arvioida käytännön toteutusta:
  - Onko tutkittava interventio riittävän hyvin määritelty ja miten varmistetaan, että se pysyy samana koko tutkimuksen ajan?
  - Toimiiko rekrytointi?
  - Onnistuvatko rekrytoinnit suunnitellussa aikataulussa?
  - Kerätäänkö prosessista riittävästi tietoa, jotta tutkimusta voidaan seurata?
  - Ovatko henkilöresurssit ja muut mahdolliset tutkimukseen liittyvät resurssit (kuten laite- ja säilytysresurssit) riittävät?
3. Kerätä tietoa satunnaistetun vertailukokeen otoskoon määrittämiseen:
  - Mitkä ovat päävasteen ja toissijaisten vasteiden kuvailevat tunnusluvut? Otokoko suositellaan laskettavaksi päävastetta käyttäen, joten esimerkiksi päävasteen hajonnasta on tärkeää saada tietoa. Näin on erityisesti silloin, kun kirjallisuudessa on hyvin vähän tietoa päävasteen jakautumisesta kohdejoukossa.
  - Onko satunnaistetulle vertailukokeelle jo laskettu otoskoko riittävä tai onko se liian suuri?

4. Tarkastella verkkosivujen tai mobiililaitteille suunniteltujen sovellusten toimivuutta ja ominaisuuksia, mikäli tutkimuksessa on tarkoitus tällaisia käyttää
- Onko tutkimushenkilöiden kirjautuminen sovellukseen tai sivustolle riittävän vaivatonta?
  - Onko sovelluksen/sivuston sisältö tarpeeksi selkeä, informatiivinen ja houkutteleva?
  - Onko tunnistettavissa jokin tietty vaihe, jossa tutkittavat usein keskeyttävät, ja voisiko tämä johtua sovelluksen tai sivuston muotoilusta ja/tai toimivuudesta?
  - Voidaanko kyseistä kohtaa muokata keskeyttävien määrän minimoimiseksi?

Ennen varsinaisen kokeen aloittamista kannattaa miettiä rekrytointinopeuden lisäksi sitä, millä tavoin ja kuinka tiiviisti tutkimuksen etenemistä voidaan seurata. Tätä varten on hyvä käyttää esimerkiksi vuokaaviota tai muuta vastaavaa prosessikuvausta, jotta mahdollisista ongelmakohdista saadaan ajantasaista tietoa. Näin säästetään tutkimusresursseja, kun mahdollisia korjausliikkeitä voidaan tehdä nopeasti.

Pilotteja voidaan tehdä useampia, ja niissä voidaan keskittyä joko yhteen tai useampaan edellä mainituista neljästä esimerkkipohdasta. Tarvittava tutkittavien määrä on riippuvainen pilotin tarkoituksesta, mutta useimmiten se on 10–50.

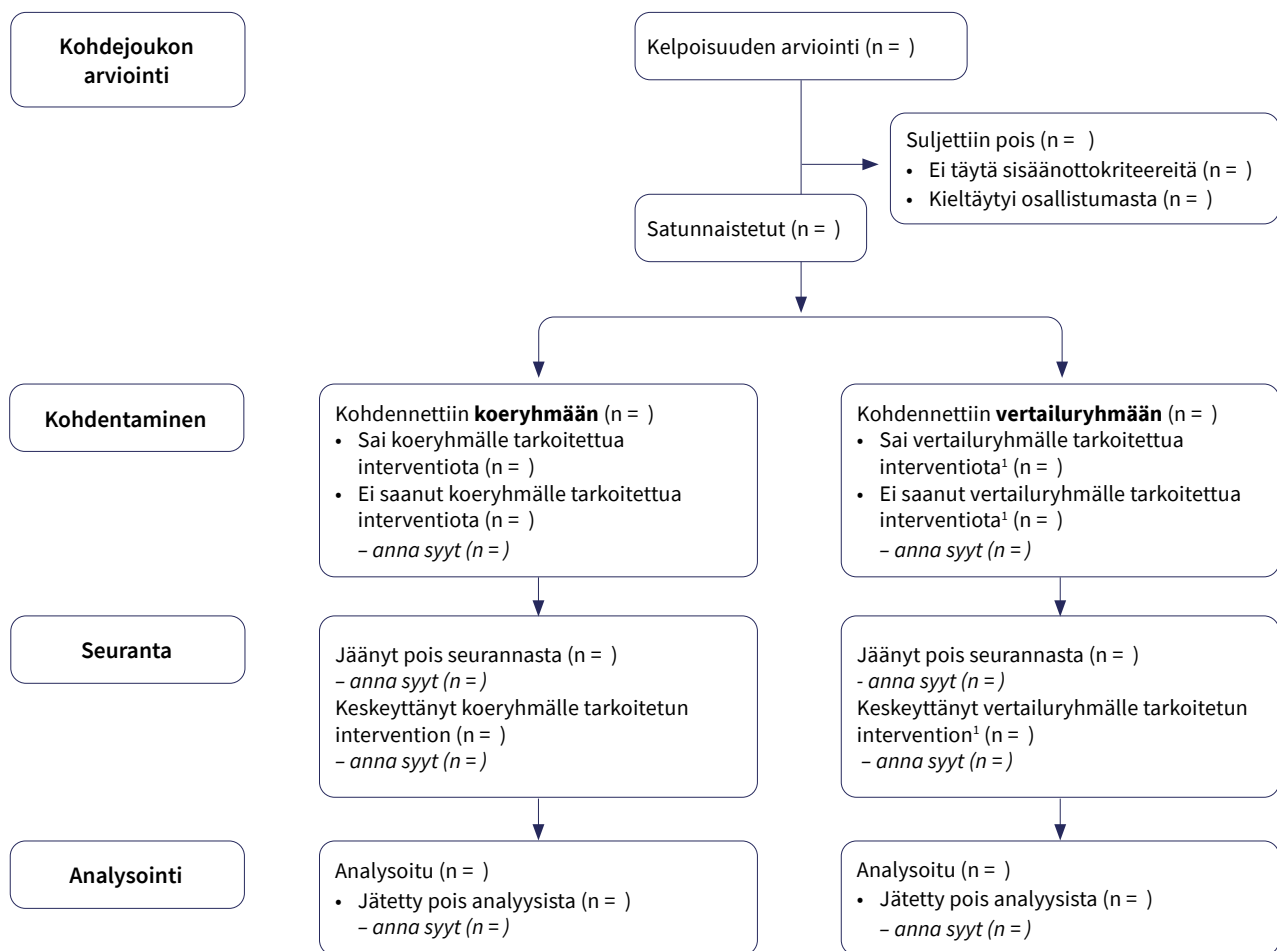
### CONSORT-RYHMÄ JA -LAUSUNTO

Vuonna 1993 joukko asiantuntijoita, joka koostui lääketieteellisten lehtien toimittajista, kliinisistä tutkijoista, epidemiologeista ja metodologeista, tapasi Ottawassa, Kanadassa. Tavoitteena oli kehittää väline satunnaistettujen vertailukokeiden raporttien laadun arvioimiseksi. Tapaaminen johti CONSORT-ryhmän (Consolidated Standards of Reporting Trials) perustamiseen ja niin sanotun CONSORT-lausunnon julkaisemiseen.

CONSORT-lausunto on väline, jota päivitetään säännöllisesti kokemuksen ja uuden empiirisen näytön perusteella. Se sisältää useita CONSORT-ryhmän kehittämää aloitteita, joiden tarkoituksena on ratkaista satunnaistettujen vertailukokeiden puutteellisesta raportoinnista aiheutuvia ongelmia. Avainjäsenet kokoontuvat säännöllisesti tarkastelemaan, onko lausuntoa tarvetta tarkentaa.

### KUVIO 1.

Satunnaistetun vertailukokeen toteuttamisen vuokaavio (CONSORT 2010).



<sup>1</sup>Vertailuryhmälle tarkoitettu interventio on useimmiten olemassa oleva standardihoito, tietopaketti tai ei mitään.

## TAULUKKO 1.

### Satunnaistetun vertailukokeen raportoinnin tarkistuslista (CONSORT 2010).

Osa-alue/Otsikko	Asia-numero	Tarkistuslistan kohta	Sivu-numero <sup>1</sup>
<b>Otsikko ja tiivistelmä</b>			
	1a	Otsikossa mainitaan satunnaistettu vertailukoe	
	1b	Jäsennely tiivistelmä koeasetelmasta, menetelmistä, tuloksista ja johtopäätöksistä	
<b>Johdanto</b>			
Tausta ja tavoitteet	2a	Tieteellinen tausta ja perustelut	
	2b	Erytetyt tavoitteet tai hypoteesit	
<b>Menetelmät</b>			
Koeasetelma	3a	Koeasetelman kuvaus (kuten rinnakkaisryhmä- tai faktorikoe) mukaan lukien koeryhmälle tarkoitettun intervention ja vertailuryhmälle tarkoitettun intervention jakosuhte (esimerkiksi 1:1, 2:1)	
	3b	Kokeen alkamisen jälkeen tapahtuneet tärkeät menetelmien muutokset ja niiden syyt (esimerkiksi muutokset kelpoisuuskriteereissä)	
Osallistujat	4a	Osallistujien kelpoisuuskrityerit	
	4b	Puitteet ja tapahtumapaikat, joissa data kerätään	
Interventiot	5	Kokeen toistamisen mahdollistava, riittävän yksityiskohtainen interventioiden määrittely; miten ja milloin interventiot todellisuudessa toteutettiin	
Tarkasteltavat muuttujat/vasteet	6a	Etukäteen määritellyt päävasteet ja toissijaiset vasteet; miten ja milloin ne todellisuudessa mitattiin	
	6b	Muutokset tarkasteltaviin vasteisiin kokeen alkamisen jälkeen ja muutosten syyt	
Otoskoko	7a	Miten otoskoko määriteltiin?	
	7b	Jos sovellettavissa, välianalyysien ja pysäytysääntöjen perustelut	
Satunnaistaminen:			
– Satunnaistamisjonojen luominen	8a	Satunnaistamisjonojen luomiseen käytetty menetelmä	
	8b	Satunnaistamisen tyyppi (kuten lohkoasetelma); käytettyjen rajoitusten yksityiskohdat	
– Mekanisimi kohdentamisen salaamiseen	9	Menetelmä, jolla satunnaistamisjonolistat tehtiin (kuten jonojen mukaan numeroidut kirjekuoret), sekä kuvaus, miten satunnaistamisen koodi salattiin interventioiden kohdentamiseen asti	
– Toteutus	10	Kuka teki satunnaistamisjonolistat, kuka rekrytoi tutkimushenkilöt mukaan tutkimukseen ja kuka toteutti teknisen satunnaistamisen?	
Sokkoutus	11a	Jos tehtiin sokkoutus, ketkä olivat sokkona intervention kohdentamisen jälkeen (esimerkiksi tutkimushenkilöt, tutkijat, statistikko) ja miten sokkoutus taattiin?	
	11b	Tarvittaessa kuvaus interventioiden samankaltaisuudesta	
Tilastolliset menetelmät	12a	Käytetyt tilastolliset menetelmät, joilla päävasteita ja toissijaisia vasteita verrattiin ryhmien välillä	
	12b	Lisäanalyysit, kuten osajoukkoanalyysit ja korjatut analyysit	
<b>Tulokset</b>			
Osallistujavirta (vuokaavio on erittäin suositeltava)	13a	Jokaisessa ryhmässä satunnaistettujen osallistujien määrä, aiottua interventiota saaneiden määrä ja analyysiin osallistuneiden määrä päävasteen osalta	
	13b	Jokaiselle ryhmälle tutkimushenkilöiden keskeyttämisten ja poissulkemisten määrät ja syyt (satunnaistamisen jälkeen)	
Rekrytointi	14a	Päivämäärät, jotka määrittelevät rekrytointi- ja seurantajakso	
	14b	Miksi tutkimus päättyi tai päätettiin	
Perustiedot	15	Taulukko, jossa esitetään kunkin ryhmän demografiset ja kliiniset perustiedot	
Analysoitavien osallistujien lukumäärä	16	Jokaisen ryhmän osalta kuhunkin analyysiin kuuluvien osallistujien määrä ja tieto siitä, tehtiinkö analyysit alkuperäisten määrättyjen ryhmien mukaan	
Vasteet ja estimointi	17a	Jokaisen päävasteen ja toissijaisen vasteen tulokset jokaiselle ryhmälle erikseen sekä arvioitu vaikutuskoko ja sen tarkkuus (kuten 95 %:n luottamusväli)	
	17b	Kaksiluokkaisille vasteille suositellaan sekä absoluuttisen että suhteellisen vaikutuskoon esittämistä	
Lisäanalyysit	18	Kaikkien muiden tehtyjen analyysien tulokset, mukaan lukien osajoukkoanalyysit ja korjatut analyysit; erotettava ennalta määriteltyjen analyysien tulokset kokeellisista	
Haitat	19	Kaikki tärkeät haittapahtumat tai ei-toivotut vaikutukset jokaisessa ryhmässä	
<b>Pohdinta</b>			
Rajoitukset	20	Tutkimuksen rajoitukset mahdollisen harhan ja epätarkkuuden vaikutukset huomioiden. Tarvittaessa: miten monivertailun ongelma huomioidaan tilastollisen merkitsevyydystason määrittämisessä?	
Yleistettävyyys	21	Tutkimuksen tulosten yleistettävyyys (ulkoinen validiteetti, sovellettavuus)	
Tulkinta	22	Tulkinta on johdonmukainen tulosten kanssa, kun hyödyt ja haitat sekä muut asiaankuuluvat seikat otetaan huomioon	
<b>Muu tieto</b>			
Rekisteröinti	23	Tutkimusrekisterin nimi ja rekisteröintinumero	
Tutkimussuunnitelma	24	Missä tutkimussuunnitelma on kokonaisuudessaan luettavissa, jos saatavilla?	
Rahoitus	25	Rahoituslähteet ja muu tuki (kuten tutkimuslääkkeet), rahoittajien rooli	

<sup>1</sup> Sivunumero viittaa tutkijan kirjoittaman tutkimusartikkelin sivuun.

## SATUNNAISTETUN VERTAILUKOKEEN TOTEUTTAMISEN VUOKAAVIO

Satunnaistettua vertailukoetta raportoitaessa mukaan on liitettävä tavallisesti kuviossa 1 esitetty CONSORT 2010 -vuokaavio sekä taulukossa 1 esitetty CONSORT 2010 -tarkistuslista täytettynä (CONSORT 2010). Eri tieteellisillä lehdillä voi olla hieman näistä poikkeavia vaatimuksia – usein kyse on lähinnä tarkentavista lisäkysymyksistä. Satunnaistettua vertailukoetta raportoitaessa suositellaan käytettävän Template for Intervention Description and Replication (TIDieR) -tarkistuslistaa CONSORT 2010 -tarkistuslistan lisäksi (Hoffman ym., 2014).

Kuviossa 1 on esitetty kaavion perusmalli kahden rinnakkaisen ryhmän satunnaistettuun tutkimukseen. Malli on helposti sovellettavissa myös tilanteeseen, jossa on useampi ryhmä. Vuokaaviossa on neljä eri vaihetta: 1) kohdejoukon arviointi, 2) kohdentaminen, 3) seuranta ja 4) analysointi. Raportoitaessa on tärkeää kirjata tutkimukseen rekrytoitujen henkilöiden määrät. Tämä koskee myös tutkimuksesta kieltäytyneitä ja henkilöitä, jotka eivät täytä sisäännotokriteereitä.

### CONSORT 2010 -TARKISTUSLISTA

Tarkistuslistassa on ohjeita kaikentyyppisten satunnaistettujen vertailukokeiden raportoimiseen, mutta erityisesti se keskittyy yleisimpiin eli yksilösatunnaistettuihin, kahden ryhmän rinnakkaisiin vertailukokeisiin. CONSORT-ryhmän julkaisemat ohjeet ovat luettavissa CONSORT-sivustolla (<http://www.consort-statement.org/>), missä on ladattavissa myös englanninkielinen tarkistuslista (word- ja pdf-muodossa).

Muut kokeet, kuten satunnaistetut klusterikokeet (cluster-randomized trials) (Campbell ym., 2012), esitutkimus- ja toteutettavuuskokeet (pilot and feasibility trials) (Eldridge ym., 2016) sekä huonommuudettomuuskokeet (non-inferiority trials) (Piaggio ym., 2012), vaativat vaihtelevan määrän lisätietoja tai muutoksia rinnakkaisryhmille tehtyyn tarkistuslistaan. Esimerkiksi satunnaistetuissa klusterikokeissa lisätiedot liittyvät niin yksilö- kuin klusteritason tuloksia esittäessä seuraaviin seikkoihin: klusterien määrittely, otoskoon laskemisessa käytetty klusterien sisäinen korrelaatiokerroin, satunnaistamisen toteuttaminen ja käytetyt tilastolliset menetelmät. CONSORT-ryhmä on julkaissut klusteri-, esitutkimus- ja toteutettavuus- sekä huonommuudettomuuskokeille erilliset tarkistuslistat, jotka ovat niin ikään ladattavissa sivustolta (<http://www.consort-statement.org/downloads/extensions>).

Tarkistuslista kattaa tutkimuksen kulun suunnittelusta raportointiin. Sitä noudattamalla tutkijan on helppo kirjoittaa laadukas tutkimusartikkeli. Tarkistuslista toimii lisäksi hyvänä työkaluna raportoinnin selkeyttä, täydellisyyttä ja läpinäkyvyyttä arvioitaessa. Selkeät kuvaukset palvelevat parhaiten lukijoiden etua.

CONSORT 2010 -lausunto ei sisällä suosituksia liittyen kokeiden suunnitteluun, toteuttamiseen tai analysointiin, vaan se koskee raportoinnista: mitä tehtiin ja mitä löydettiin. Siitä huolimatta tarkistuslista vaikuttaa epäsuorasti kokeiden suunnitteluun ja toteuttamiseen. Läpinäkyvä raportointi paljastaa tutkimuksen mahdolliset puutteet. Jos laadultaan riittämättömiä kokeita sisältävä tutkimus raportoidaan avoimesti, sen ei pitäisi pystyä läpäisemään julkaisuprosessia ilman, että tutkimuksen puutteet tulevat esiin.

Tarkistuslistan tavoitteena onkin ohjata tutkijoita kokeiden parempaan suunnitteluun, toteutukseen ja raportointiin. Lista koostuu kuudesta osa-alueesta: 1) otsikko ja tiivistelmä, 2) johdanto, 3) menetelmät, 4) tulokset, 5) pohdinta ja 6) muu tieto. Seuraavassa keskitytään niihin

CONSORT 2010 -tarkistuslistan kohtiin, jotka usein jäävät huomioimatta ensimmäistä kertaa satunnaistettua vertailukoetta raportoitaessa. Suluissa olevat numerot viittaavat taulukon 1 kohtiin 1–25.

### Otsikko ja tiivistelmä

Jotta satunnaistetusta vertailukokeesta raportoiva artikkeli olisi tunnistettavissa sähköisestä tietokannasta, sen otsikon tulee kertoa, millaisesta tutkimuksesta on kyse (Moher ym., 2010). Otsikossa kannattaa käyttää ”satunnaistettu”-sanaa osoittamaan, että osallistujat jaettiin satunnaisesti ryhmiin (kohta 1a). Ohjeen mukainen otsikko on esimerkiksi ”The Effectiveness of the Bullying Prevention Program in Finland: Results from a Randomized Controlled Trial”.

### Johdanto

Tavoitteet (Objectives) ovat kysymyksiä, joihin kokeen oli tarkoitus vastata. Ne liittyvät esimerkiksi tietyn terapeuttisen tai ennaltaehkäisevän intervention tehokkuuteen. Hypoteesit ovat ennalta määritettyjä kysymyksiä, joita testataan tavoitteiden saavuttamiseksi. Hypoteesit ovat tarkempia kuin tavoitteet, ja niitä voidaan arvioida tilastollisesti. Useimmat satunnaistetuista vertailukokeista tehdyt julkaisut raportoivat tavoitteet ja hypoteesit riittävän selkeästi (kohta 2b).

### Menetelmät

Kohdassa 3a (koeasetelman kuvaus) ”design”-sanaa käytetään usein laajassa merkityksessä viittaamaan kaikkiin koejärjestelyihin liittyviin asioihin, mukaan lukien satunnaistaminen ja sokkoutus. Kaksi viimeksi mainittua liittyvät kuitenkin tutkimuksen tyyppiin eivätkä koeasetelmaan. Yleisin koeasetelma on rinnakkaisten ryhmien koeasetelma, mutta myös ristikkäiskaavio- ja faktorikokeet ovat mahdollisia. Näiden laajempien kokeiden suunnittelun erityisnäkökohtia, kuten satunnaistamisen ja sokkoutuksen yksityiskohtia, käsitellään kohdissa 8–11.

CONSORT-lausunto keskittyy kuitenkin pääasiassa kokeisiin, joissa osallistujat satunnaistetaan yksittäin toiseen kahdesta ”rinnakkaisesta” ryhmästä. Itse asiassa hieman yli puolessa maailmalla julkaistuista satunnaisista vertailukokeista on käytetty rinnakkaisryhmä-asetelmaa (Chan & Altman, 2005).

*”Tarkistuslistan tavoitteena on ohjata tutkijoita kokeiden parempaan suunnitteluun, toteutukseen ja raportointiin.”*

Kokeen osallistujien valinnassa käytetyistä kelpoisuuskriteereistä tarvitaan kattava kuvaus. Kriteerien ymmärtäminen on tärkeää, jotta voidaan arvioida, keitä tutkimuksen tulokset koskevat. Tutkimuksen yleistettävyyttä (sovellettavuutta) ja merkitystä kliinisen tai kansanterveyden kannalta on mahdollista arvioida vain, kun tutkimusjoukko on ymmärrettävästi määritelty (kohta 4a) (Moher ym., 2010).

Tutkijan on kuvattava koeryhmälle ja vertailuryhmälle annetut interventiot yksityiskohtaisesti (kohta 5). Esimerkiksi jos kyse on terapiasta, tulee kuvata, kuinka usein ja pitkään terapiaa on annettu. Jos verrokkiryhmä saa ”tavallista hoitoa” (treatment as usual), on tärkeää kuvata, mitä se tarkoittaa. Jos vertailuryhmä tai interventioiryhmä saa interventioiden yhdistelmän, tutkijan on kuvattava jokainen interventio erikseen. Samoin tulee kuvata järjestys, jossa

yhdistelmän interventiot otetaan käyttöön tai lopetetaan, ja tarvittaessa niiden käyttöönoton alkamiseen vaikuttavat tekijät.

Tutkijan on varmistettava, etteivät koeryhmälle tai vertailuryhmälle annettavat interventiot muutu tutkimuksen aikana. Tämä on erityisen tärkeää pitkäkestoisissa (kuukausien tai vuosien pituisissa) tutkimuksissa, joissa interventiot ovat terapiamuotoisia. Varmistumenetelmät on hyvä raportoida riittävän yksityiskohtaisesti. Mikäli interventiot ovat esimerkiksi tablettimuotoisia lääkehoitoja tai täysin strukturoituja internet-pohjaisia ohjelmia, niiden muuttumattomuus on huomattavasti helpompi varmistaa.

Päävaste (primary outcome) määritellään ennalta tutkimussuunnitelmassa. Sen katsotaan olevan tutkimuksen tärkein vaste asianomaisille sidosryhmille (kuten kohderyhmälle tai potilaille, poliittisille päättäjille, klinikoille tai rahoittajille), ja sitä käytetään yleensä otoskoon laskennassa. Päävasteen tulokset on ilmoitettava selkeästi sellaisinaan satunnaistetun vertailukokeen raportissa (European Medicines Agency, 1998; Moher ym., 2010). Muut kiinnostavat vasteet ja tulokset ovat toissijaisia tai muita vasteita (secondary outcomes, additional outcomes) (kohta 6a).

Kohdan 6b mukaan kirjoittajien tulee raportoida kaikista merkittävistä muutoksista protokollaan. Tällaisia ovat myös kelpoisuus-kriteereihin, interventioihin, tutkimuksiin, tiedonkeruuseen, analyysimenetelmiin ja tuloksiin liittyvät odottamattomat muutokset. Tutkimuksen otoskoko on tieteellisistä ja eettisistä syistä suunniteltava huolellisesti siten, että lääke-, hoito- ja tilastotieteelliset näkökohdat ovat tasapainossa (Hinkka-Yli-Salomäki, 2021).

Ilhanetapauksessa tutkimuksen otoskoon tulisi olla riittävän suuri, jotta todennäköisyys (voima) havaita tietyn kokoinen kliinisesti tärkeä ero tilastollisesti merkitsevänä on riittävä – jos tällainen ero on olemassa. Tärkeänä pidetyn vaikutuksen koko on kääntäen verrannollinen sen havaitsemiseen tarvittavaan otoskoko; pien-ten erojen havaitsemiseksi tarvitaan suuria otoskokoja.

Otetaan esimerkiksi tilanne, jossa tutkija pyrkii osoittamaan uuden masennusintervention tehon olemassa olevaan tavalliseen hoitoon verrattuna ja päävasteena on Edinburgh Depression Scale (EPDS) -kysely. Tutkija arvioi, että pisteen ero yhteispistemäärässä

## ”Otsikon tulee kertoa, millaisesta tutkimuksesta on kyse.”

on jo kliinisesti merkittävä hoitojen välinen ero. Tällöin kirjallisuudesta poimittua keskimääräistä hajontaa (SD) käyttäen tarvittavaksi otoskooksi saadaan 400 tutkittavaa molemmissa ryhmissä. Tarvittaisiin siis yhteensä 800 tutkimushenkilön EPDS-tiedot, jotta havaittaisiin tilastollisesti merkitsevä ero uuden masennusintervention ja tavallisen hoidon välillä, mikäli ryhmien välillä on todellisuudessa eroa. Jos taas tutkija arvioisi, että uuden masennusintervention vaikutus EPDS-kyselyn tulokseen on keskimäärin 3 pistettä suurempi käypään hoitoon verrattuna, tarvittavaksi otoskooksi riittäisi 100 tutkittavaa kummassakin ryhmässä.

Otoskoon laskemiseen tarvitaan:

- 1) arvioidut tulokset kussakin ryhmässä (kliinisesti tärkeä tavoite-ero interventio- ja vertailuryhmien välillä)
- 2)  $\alpha$  eli tyypin I -virhetaso (tavallisesti 0,05)
- 3) tilastollinen voima  $[1-\beta]$  ( $\beta$  = tyypin II -virhetaso), tyypillisesti 0,80–0,95, mikä ilmoitetaan usein prosentteina eli 80–95 %)
- 4) jatkuvien tulosten osalta mittauksen keskihajonnat (7a).

Tarkistuslistan kohdat 8–10 käsittelevät satunnaistamista. Satunnaistamisprosessissa on kolme vaihetta: satunnaistamisjonojen luominen (sequence generation), kohdentamisen salaaminen (allocation concealment) ja toteutus (implementation). Samat henkilöt voivat suorittaa useamman kuin yhden vaiheen, mutta satunnaistamisjonojen tekemiseen ja kohdentamisen salaamiseen osallistuvien tulisi olla eri henkilöitä kuin tutkimuksen toteuttamiseen osallistuvien. Suositeltavaa siis on, että henkilöt, jotka ovat mukana satunnaistamisprosessin kahdessa ensimmäisessä vaiheessa eivät ole mukana toteutusvaiheessa.

## ”Kokeen osallistujien valinnassa käytetyistä kelpoisuus-kriteereistä tarvitaan kattava kuvaus.”

Itse olen statistikkona tehnyt satunnaistamislistoja ja kehittänyt kohdentamisen salaamisen prosesseja lukuisiin satunnaistettuihin vertailukokeisiin, mutten ole ollut suorassa yhteydessä tutkittaviin datan keräämisvaiheessa, eli en ole osallistunut tutkimuksen toteuttamiseen. Teknisen satunnaistamisen eli satunnaistamis-koodin liittämisen tutkittavaan voi tehdä esimerkiksi data manager -tyyppisessä tehtävässä oleva henkilö, jolla ei tarvitse olla satunnaistamisjonoista kuin kooditason tiedot (esimerkiksi A, B). Valinnasta johtuvan harhan minimoimiseksi satunnaistamiskoodien merkitykset (eli esimerkiksi A = vertailuryhmälle annettava interventio ja B = koeryhmälle annettava interventio) tuntevien henkilöiden määrää kannattaa pitää mahdollisimman pienenä.

Hyvä periaate on kuvata tilastolliset menetelmät riittävän yksityiskohtaisesti, jotta asiantunteva lukija, jolla on pääsy alkuperäisiin aineistoihin, voi tarkistaa tulosten oikeellisuuden.

### Tulokset

Seuraavaksi tarkastellaan CONSORT 2010 -tarkistuslistan Tulokset-osiota. Kuvion 1 osallistujavirtaa kuvaavien lukujen on oltava yhtenevät analyyseissä käytettävien aineistojen tutkimushenkilöiden lukumäärien kanssa. Yleisesti ottaen, mikäli on kyse rinnakkaisten ryhmien vertailukokeesta, tulokset esitetään siten, että mukana ovat kaikki satunnaistetut tutkimushenkilöt ITT-periaatteen mukaisesti (intention-to-treat principle).

Esimerkiksi jos viisi tutkimushenkilöä keskeytti tutkimuksen ennen päävasteena olevan kyselyn täyttämistä, ITT-periaatteen mukaan heidät on kuitenkin otettava mukaan analysoitavaan aineistoon. Toisena esimerkkinä on tilanne, jossa yksi koeryhmään satunnaistetuista tutkimushenkilöistä päätyykin inhimillisen virheen vuoksi saamaan vertailuryhmälle tarkoitettua interventiota. ITT-periaatteen mukaan kyseinen henkilö otetaan mukaan analysoitavaan aineistoon alkuperäisen satunnaistamisjärjestyksen perusteella, eli hän on mukana koe- eikä vertailuryhmän tuloksissa.

Eriyistäpaauksissa tutkittavia voidaan poistaa analysoitavasta aineistosta harhan mahdollisuutta lisäämättä.

Tällöin seuraavien ehtojen tulee täytyä:

- 1) tutkimuksen sisäänottokriteeri mitattiin ennen satunnaistamista
- 2) tutkimuksen kelpoisuusrikkomusten (eligibility violations) havaitseminen voidaan tehdä täysin objektiivisesti
- 3) kaikki henkilöt saavat yhtäläisen tarkastelun kelpoisuusrikkomusten osalta
- 4) kaikki havaitut tietyn osallistumiskriteerin rikkomukset suljetaan pois (European Medicines Agency, 1998).

Joidenkin satunnaistettujen vertailukokeiden suunnittelu ja toteuttaminen on yksinkertaista, ja osallistujavirtaa voidaan kuvata riittävästi muutamalla lauseella. Näin on varsinkin silloin, kun seurannassa ei tule menetyksiä (lost to follow-up) tai poissulkemisia (exclusions). Monimutkaisemmissa tutkimuksissa lukijoiden voi olla vaikeaa havaita, saivatko kaikki osallistajat heille osoitettua interventiota, menetettiinkö osa seurattavista tai jätettiinkö osa heistä analyysin ulkopuolelle tutkimusryhmän toimesta.

## *”Pienten erojen havaitsemiseksi tarvitaan suuria otoskokoja.”*

Nämä tiedot (määrät ja syyt) ovat tärkeitä useista syistä. Osallistajat, jotka suljetaan pois satunnaistamisen jälkeen, eivät todennäköisesti edusta kaikkia tutkimuksen osallistujia (Moher ym., 2010). Seurantavaiheessa ”luonnollisesti” tapahtuvista menetyksistä johtuva, usein väistämätön, tutkimusjoukon pieneneminen on erotettava tutkijan tekemästä aktiivisesta poissulkemisestä. Aktiivinen poissulkeminen voi liittyä esimerkiksi tutkimushenkilöiden kelpoisuuteen, hoidosta vetäytymiseen tai tutkimussuunnitelman puutteelliseen noudattamiseen. Jos tutkittavia jätetään analyysin ulkopuolelle ilman pätevää syytä, voidaan päätyä tekemään virheellisiä johtopäätöksiä. Näin voi tapahtua erityisesti, jos poissulkemiset keskittyvät joko interventio- tai vertailuryhmään.

Raportin tuloksiin sisällytetään päivämäärät, jotka määrittelevät rekrytointi- ja seurantajaksot (kohta 14a). Kun tiedetään, milloin tutkimus tehtiin ja milloin osallistujia rekrytoitiin, tutkimus asetuu historialliseen kontekstiin. Kelpoisuuskriteerit määrittävät, ketkä ovat sopivia osallistumaan kokeeseen, mutta on myös tärkeää saada olennaiset tiedot todellisuudessa mukana olleiden osallistujien ominaisuuksista. Satunnaistettujen vertailukokeiden tarkoituksena on vertailla osallistujaryhmiä, jotka eroavat toisistaan vain interventioiden suhteen. Oikein tehty satunnaistaminen estää valintaharhan, mutta se ei takaa, että ryhmät ovat lähtökohtaisesti samanlaisia. Tällöin mahdolliset erot ryhmien osallistujien ominaisuuksissa lähtötilanteessa ovat kuitenkin seurauksena sattumasta eivätkä harhaanjohtamisesta. Tutkittavia ryhmiä tulisi verrata tärkeiden demografisten ja muiden ominaisuuksien suhteen lähtötilanteessa, jotta lukijat voivat arvioida, kuinka samanlaisia ryhmät olivat. Myös vasteiden perustasot (ns. baseline-mittaus) tulee mitata ennen satunnaistamista.

CONSORT 2010 -tarkistuslistan kohta 16 ”jokaisen ryhmän osalta kuhunkin analyysiin kuuluvien osallistujien määrä, ja tieto siitä, että tehtiinkö analyysit alkuperäisten määrättyjen ryhmien mukaan” on olennaista raportoida kunkin vasteen osalta. Erityisesti kyselytutkimuksissa lukumäärät voivat olla interventio- ja vertailuryhmissä eri vasteille erilaisia. Osallistajat eivät aina saa täydellistä tai koko interventiota, ja joskus tutkimukseen kelpaamattomia (ineligible) tutkittavia on voitu satunnaisesti kohdentaa ryhmään virheellisesti. Suositeltu tapa käsitellä tällaisia ongelmia on analysoida kaikki tutkittavat heidän alkuperäisen satunnaistamiskoodinsa mukaisesti riippumatta siitä, mitä myöhemmin on tapahtunut.

Aiemmin esitettyä analysointistrategiaa ”kaikki satunnaistetut tutkimushenkilöt ITT-periaatteen mukaisesti mukaan” ei siis ole aina yksinkertaista toteuttaa. Tutkimustulokset tulee raportoida yhteenvetona kunkin ryhmän tuloksista jokaisen vasteen osalta (esimerkiksi tutkittavien määrä yhdessä tutkittavan tapahtuman kanssa tai ilman sitä; numeerisen vasteen tapauksessa keskiarvo ja

keskihajonta) sekä ryhmien välisestä vaikutuksen suuruuden erosta (effect size).

Vaikutuksen suuruuden laskeminen riippuu vasteen tyypistä:

1. Kaksiluokkaisissa vasteissa vaikutuksen suuruus voidaan esittää riskisuhteena (suhteellinen riski), vetosuhteena (odds ratio, OR) tai riskien välisenä erona (risk difference). Kohtuullisen yleisissä sairauksissa suhteellinen riski lähellä yhtä saattaa viitata kliinisesti merkittäviin eroihin kansanterveyden kannalta. Sitä vastoin harvinaisissa sairauksissa suuri suhteellinen riski ei välttämättä ole yhtä tärkeä kansanterveyden kannalta (vaikka se voi olla tärkeä niiden henkilöiden kannalta, jotka kuuluvat suuren riskin ryhmään).
2. Tarkasteltaessa tutkimushenkilöiden eloonjäämisaika-eroja, eli aikaväliä tutkimuksen aloittamisesta (tavallisesti satunnaistaminen) päätetapahtumaan, vaikutuksen suuruutta kuvaa vaarasuhde (hazard ratio) tai ero mediaanieloonjäämisajan välillä. Analysointimenetelmä on nimeltään elinaika-analyysi ja päätetapahtuma voi olla esimerkiksi sairaudesta parantuminen tai sairauden uusiutuminen.
3. Jatkuviissa vasteissa vaikutuksen suuruus voidaan esittää ryhmien välisenä keskiarvojen (tai hyvin vinojen jakaumien kohdalla mediaanien) erona.
4. Ryhmien välisten vertailujen luottamusväliä tulee esittää kaikissa vastetyypeissä.

Yleinen virhe on esittää kunkin ryhmän tulokselle erilliset luottamusväliä hoidon vaikutuksen sijaan. Otetaan esimerkiksi kahden rinnakkaisryhmän tutkimus, jossa koeryhmän interventiona on internet-pohjainen ohjausohjelma aggressiivisen käytöksen hallitsemiseksi. Vertailuryhmä jää ilman hoitoa. Vertailtaessa aggressiivisen käytöksen hallintaa koeryhmässä ennen ja jälkeen intervention saadun tulos, että interventio on tilastollisesti merkittävästi parantanut aggressiivisuuden hallintaa [keskiarvo 1,8 (SD 0,2), p-arvo 0,045, ja alkumittauksen ja loppumittauksen välisen eron luottamusväli 1,2–2,3].

## *”Tilastollisen merkitsevyyden ja kliinisen merkityksen välinen ero on hyvä pitää mielessä.”*

Tämä ei vielä kerro intervention vaikutuksesta, sillä saman suuntainen aggressiivisuuden hallinnan parantuminen on voinut tapahtua myös vertailuryhmässä. Tutkijan on esitettävä ryhmien välinen vertailu esimerkiksi vähentämällä interventio-ryhmän loppumittauksen keskiarvosta vertailuryhmän loppumittauksen keskiarvo. Keskiarvojen erolle lasketaan 95 %:n luottamusväli. Jos saatu luottamusväli sisältää keskiarvoja vertailtaessa luvun 0, saatu tulos ei ole tilastollisesti merkitsevä. Tämä tarkoittaa sitä, että tutkimus ei tue intervention vaikuttavuutta verrattuna kontrolliin.

Jos ryhmien välillä ei havaita tilastollisesti merkitsevää eroa, luottamusväli saattaa antaa tukea esimerkiksi kliiniselle vaikuttavuudelle. Tuloksia ei pitäisi raportoida pelkästään p-arvoina, mutta p-arvot voidaan raportoida luottamusvälien lisäksi. Tulokset tulee



raportoida kaikkien suunniteltujen päävasteiden ja toissijaisten vasteiden osalta – ei vain niiden, jotka olivat tilastollisesti merkitseviä. Valikoiva raportointi tutkimuksissa on laajalle levinnyt ja tiedostettu ongelma (Chan ym. 2004).

Tutkimussuunnitelmaan kirjattujen päävasteiden sekä toissijaisten ja muiden vasteiden analyysien lisäksi julkaisussa on selkeästi ilmoitettava muut tehdyt lisäanalyysit. Tällaisia ovat esimerkiksi osajoukkoanalyysit ja vakioidut analyysit (adjusted analyses) (kohta 18). Useiden analyysien tekeminen samasta tiedosta aiheuttaa väärin positiivisten löydösten riskin (multiplicity problem), mitä on syytä välttää. Tutkimussuunnitelmassa ennalta määritellyt analyysit ovat huomattavasti luotettavampia kuin jälkikäteen aineiston perusteella nousseet analyysit. Tämän vuoksi tutkijan tulee rekisteröidä satunnaistettu vertailukoe johonkin tutkimusrekisteriin jo ennen tutkimuksen aloittamista (esimerkiksi ClinicalTrials.gov) (Hinkka-Yli-Salomäki, 2021).

## **”Tutkimussuunnitelman ennakkorekisteröinti pienentää todennäköisyyttä, että muutoksia tehdään tutkimuksen jo käynnistettyä.”**

Tutkimusrekisterissä listataan tutkimusten vähimmäisominaisuudet, kuten käytettävä koeasetelma, ryhmien kuvaukset, pää- ja toissijaiset vasteet sekä suunnitellut tutkimusryhmien koot. Mikäli osajoukkoanalyysit tehdään, tutkijan tulee raportoida, mitä alaryhmiä tutkittiin ja miksi, ja mitkä analyysit oli ennalta määritettyjä. Osajoukkoanalyysien valikoiva raportointi voi johtaa suuremman harhaan.

Sama suositus koskee myös vakioituja analyysit, joissa vasteiden tuloksia vakioidaan lähtötason tiedoilla tai demografisilla tiedoilla. Jos näitä selittäviä muuttujia (kovariaatteja) käytetään tilastollisissa malleissa, on raportoitava sekä vakioimattomat (unadjusted) että vakioidut (adjusted) tulokset. Tutkimussuunnitelmassa tulee ilmoittaa, tehdäänkö vastemuuttujille vakiointi muuttujan lähtötason arvoilla kovarianssianalyysin avulla. Jos vastemuuttujan arvot eroavat lähtötasolla ja tehdään kovarianssianalyysi, joka korjaa lähtötason eroa, intervention arvioitu vaikuttavuus on todennäköisesti vääristynyt (Moher ym., 2010).

Tilastollisen merkitsevyyden ja kliinisen merkityksen välinen ero on hyvä pitää mielessä. Tutkijan tulee erityisesti välttää sitä yleistä virhettä, että tulkitaan tilastollisesti ei-merkitsevä tulos osoituksena tutkittavan interventio- ja vertailuryhmän vastaavuudesta (Hinkka-Yli-Salomäki, 2021). Interventio- ja vertailuryhmän vastaavuus jonkin vasteen suhteen voidaan todentaa päättämällä etukäteen sallittu ala- ja yläraja eron luottamusvälille ja katsomalla jälkikäteen, sijoittuuko tutkimuksesta laskettu 95 %:n luottamusväli sallitun ala- ja ylärajan välille.

Ei-merkitsevä tulos voi johtua myös liian pienestä otoskoota. Todellisuudessa ryhmien välillä voi olla ero tutkittavassa vasteessa, mutta tutkittavien määrä on liian pieni osoittamaan sitä. Erityisen varovainen tulee olla arvioitaessa yksittäisten vertailujen tuloksia monivertailutilanteessa. Monivertailutilanne voi johtua esimerkiksi useista rinnakkaisryhmistä, päävasteista, aikapisteistä tai osajoukkoanalyysistä. Tällaisissa tapauksissa osa tuloksista voi olla tilastollisesti merkitseviä jo pelkästään sattumalta.

## **Pohdinta**

CONSORT 2010 -tarkistuslistan Pohdinta-osiossa tutkijaa pyydetään pohtimaan tutkimuksen rajoituksia, mahdollisia harhan lähteitä ja monivertailuongelmaa (kohta 20).

Tutkimustulosten ulkoista validiteettiä eli sitä, missä määrin tulokset voidaan yleistää muihin olosuhteisiin, kutsutaan myös yleistettävyydeksi tai sovellettavuudeksi. Ulkoisen validiteetin edellytyksenä on sisäinen validiteetti eli se, että tutkimuksen suunnittelu ja toteutus auttavat poistamaan harhan mahdollisuuden (kohta 21). Tutkijalle esitetään pohdittavaksi esimerkiksi seuraavat kysymykset:

- Voidaanko tulokset yleistää koskemaan yksittäisiä henkilöitä tai ryhmiä, jotka poikkeavat tutkimukseen osallistuneista esimerkiksi iän, sukupuolen, sairauden vaikeusasteen tai muiden sairauksien osalta?
- Voidaanko samanlaisia tuloksia odottaa esimerkiksi perus- ja erikoissairaanhoidon tasolla?
- Mikä on niiden tekijöiden vaikutus vasteisiin, joita tutkimuksessa ei arvioitu tai ei pystytty arvioimaan?
- Millainen merkitys on seurannan pituudella ja hoidon/hoidonohjelman kestolla?

Ulkoisen validiteetin toteutuminen on aina harkittava tapauskohtaisesti. Se on riippuvainen tutkittavien ominaisuuksista, tutkimusympäristöstä, interventiosta ja valitusta vertailuryhmästä sekä vasteista. Siksi on erittäin tärkeää raportoida kelpoisuus-kriteerit ja tutkimusympäristö sekä kuvata annetut interventiot ja verrokkit (mitä interventiota mikin ryhmä kussakin vaiheessa sai).

Satunnaistetun vertailukokeen raportissa tulee ilmoittaa rekisterin nimi ja numero (kohta 23), esimerkiksi näin: ”Tutkimus on rekisteröity osoitteessa ClinicalTrials.gov, numero NCT00244842.” Tutkimussuunnitelma on tärkeä kirjoittaa kokonaisuudessaan, koska rekisteröinnissä kysyttävä tieto kattaa vain osan satunnaistetun vertailukokeen ominaisuuksista (mm. otoskoko, tutkimuspaikka, pää- ja toissijaiset vasteet, tutkittava interventio) (kohta 24).

Kokonaisten tutkimussuunnitelmien saattaminen lukijoiden saataville on tärkeää tutkimuksen laadun kannalta, ja sille on useita tapoja. Jotkut aikakauslehdet (kuten Trials) julkaisevat myös pelkkiä tutkimussuunnitelmia. Tällaiseen julkaisuun on helppo viitata, kun raportoidaan tutkimuksen tärkeimmistä tuloksista. Tämän lisäksi tutkimustuloksia julkaisevat lehdet voivat laittaa tuloksista raportoitavan artikkelin yhteyteen verkkosivulleen kyseisen tutkimuksen suunnitelman. Tutkimussuunnitelman ja -tulosten saatavuus paranevat, kun käytetään avointa julkaisufuorua (open access).

## **”Ei-satunnaistetuissa tutkimuksissa saadaan yleensä suurempia arvioituja hoitovaikutuksia kuin satunnaistetuissa.”**

Sekä tutkimuksen rekisteröinti että tutkimussuunnitelman julkaiseminen pienentävät todennäköisyyttä, että muutoksia tehdään tutkimuksen jo käynnistettyä, ja ehkäisevät toivottavasti myös valikoivaa raportointia. Satunnaistetun vertailukokeen rekisteröinti varmistaa myös, että monet tutkimusprotokollan tiedot ovat saatavilla tietokannasta. Tutkijan on ilmoitettava myös tutkimuksen rahoituslähteet, sillä tämä on tärkeää tietoa arvioinnin kannalta (25). On osoitettu, että ainakin lääketieteellisuuden rahoittama tutkimus tuottaa todennäköisemmin tutkimusta sponsoroivan yrityksen tuotetta suosivia tuloksia kuin muista lähteistä rahoitetut tutkimukset (Moher ym., 2010).

Hyvinvointia lisäävien toimenpiteiden arviointi voi olla harhaanjohtavaa, elleivät tutkijat varmistaa, että vertailut ovat mahdollisimman harhattomia ja puolueettomia. Tutkittavien satunnaistaminen tutkittaviin ryhmiin on ainoa menetelmä, joka poistaa valikoitumisesta ja mahdollisista muista muuttujista johtuvat harhat. Ei-satunnaistetuissa tutkimuksissa saadaan yleensä suurempia arvioituja hoitovaikutuksia kuin satunnaistetuissa tutkimuksissa.

Myös huonosti toteutetut satunnaistetut vertailukokeet ovat alttiita tutkimustulosten luotettavuuden vaarantaville harhoille. Tutkimukset, joissa on käytetty riittämättömästi tai epäselvästi kuvattua menetelmää satunnaistamiskoodin salaamiseksi, tuottivat 31 % suuremmat arvot vaikutuksesta kuin ne, joissa on käytetty riittävää menetelmää koodin salaamiseksi (Wood ym., 2008). Vastaavasti avoimet satun-

naistetut vertailukokeet (open studies), tuottivat 25 % suurempia arvioita hoidon vaikuttavuudesta (Wood ym., 2008).

## LOPUKSI

Satunnaistetun vertailukokeen suunnittelu ja toteuttaminen edellyttävät metodologista ja kliinistä asiantuntemusta, huolellista työtä ja valppautta odottamattomien vaikeuksien varalta. Raportti on kirjoitettava huolella: menetelmäkuvausten pitää olla riittävän yksityiskohtaisia ja läpinäkyviä, jotta lukijat voivat vakuuttua, että harhojen mahdollisuus on minimoitu ja tulokset ovat mahdollisimman puolueettomia. Laadukas raportointi on osa laadukasta tiedettä. ●

## AVAINSANAT:

vaikuttavuus, RCT, satunnaistettu vertailukoe, pilotointi, raportointi

## SIDONNAISUUDET:

Sanna Hinkka-Yli-Salomäki: Ei sidonnaisuuksia.

## LÄHTEET

Campbell, M. K., Piaggio, G., Elbourne, D. R. & Altman D. G. (2012). Consort 2010 statement: extension to cluster randomised trials. *BMJ*, 345:e5661. <https://doi.org/10.1136/bmj.e5661>

Chan, A. W. & Altman, D. G. (2005). Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet*, 365:1159-62. [https://doi.org/10.1016/S0140-6736\(05\)71879-1](https://doi.org/10.1016/S0140-6736(05)71879-1)

Chan, A. W., Hróbjartsson, A., Haahr, M. T., Gøtzsche, P. C. & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*, 291:2457-65. <https://doi.org/10.1001/jama.291.20.2457>

CONSORT 2010. Download the CONSORT 2010 publications. CONSORT Transparent Reporting of Trials. Haettu 19.2.2022 osoitteesta <http://www.consort-statement.org/downloads/consort-statement>

CONSORT. History. CONSORT Transparent Reporting of Trials. Haettu 19.2.2022 osoitteesta <http://www.consort-statement.org/about-consort/history>

Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., ... & on behalf of the PAFS consensus group. (2016). CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *Pilot Feasibility Stud* 2, 64. <https://doi.org/10.1186/s40814-016-0105-8>

European Medicines Agency. (1998). ICH E9 statistical principles for clinical trials. ICH Harmonised Tripartite Guideline. <http://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials>

Hinkka-Yli-Salomäki, S. (2021). Vaikuttavuustutkimus: Satunnaistettu vertailukoe tulee suunnitella huolella. *Kasvun tuki*. Haettu 19.2.2022 osoitteesta <https://kasvuntuki.fi/aikakauslehti/julkaisut/kasvun-tuki-aikakauslehti-2021/hinkka-yli-salomaki/>

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., ... & Michie, S. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, 348:g1687. <https://doi.org/10.1136/bmj.g1687>

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>

Piaggio, G., Elbourne, D. R., Pocock, S. J., Evans, S. J. W., Altman, D. G., for the CONSORT Group. (2012). Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*, 308(24): 2594-2604. <https://doi.org/10.1001/jama.2012.87802>

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G., ... & Sterne, J. A. C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*, 2008;336:601-5. <https://doi.org/10.1136/bmj.39465.451748.AD>