

## Vaikuttavuustutkimus:

# Miten arvioida harhaa satunnaistetuissa vertailukokeissa?

Artikkeli on jatkoa numerossa 2/2021 julkaistulle artikkelille Vaikuttavuustutkimus: Satunnaistettu vertailukoe tulee suunnitella huolella ja numerossa 1/2022 julkaistulle artikkelille Vaikuttavuustutkimus: Satunnaistetun vertailukokeen laadukkaan raportoinnin edellytykset.

- Satunnaistamattomien vertailukokeiden harha johtaa usein uusien interventioiden vaikutusten yliarviointiin.
- Satunnaistetun vertailukoeasetelman käyttö ei takaa tutkimuksen ja siitä kirjoitetun artikkelin luotettavuutta ja laatua.
- Yleisimmät harhan lähteet voidaan luokitella viiteen luokkaan: 1) valinta-, 2) toiminta-, 3) hiertymis-, 4) havaitsemis- ja 5) raportointiharha.
- Harhalla on suora vaikutus tutkimustulosten luotettavuuteen.
- Cochranen harhan riski -työkalulla voidaan arvioida artikkelin luotettavuutta ja laatua.



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

**M**ielenterveyden tukemiseen tarkoitettujen psykososiaalisten menetelmien – aivan kuten muidenkin interventioiden – vaikuttavuuden osoittamisen kultainen standardi on satunnaistettu vertailukoe. Satunnaistettujen vertailukokeiden malleja on lukuisia ja ne voivat olla hyvinkin erilaisia esimerkiksi koeasetelmaltaan ja tilastolliselta metodologialtaan. Tämä luo haasteen lukijalle: miten lukija pystyy artikkelin perusteella tekemään oman arvionsa tulosten luotettavuudesta, merkitsevyydestä ja yleistettävyydestä, kun olemassa on useita satunnaistetun vertailukokeen malleja? Mihin mahdollisiin harhan lähteisiin lukijan pitäisi kiinnittää huomiota? Miten lukija voi vakuuttua, että artikkeli on riittävän laadukas, jotta sen tulokset ovat yleistettävissä? On tärkeää pitää mielessä, että satunnaistetun vertailukoeasetelman käyttö ei takaa tutkimuksen ja siitä kirjoitetun artikkelin luotettavuutta ja laatua.

### MITÄ ON HARHA?

Harha kuvataan tavallisesti mittaustuloksen järjestelmälliseksi taipumukseksi poiketa tiettyyn suuntaan todellisesta arvosta. Järjestel-

mällinen poikkeaminen todellisesta arvosta voi johtaa intervention vaikutusten yli- tai aliarviointiin. Yleensä olemme kiinnostuneita osoittamaan, että koeryhmälle annettu interventio on vertailuryhmälle annettua interventiota parempi, tehokkaampi ja turvallisempi.

Uuden intervention vaikuttavuutta saatetaan yliarvioida, jos sitä on tutkittu vain satunnaistamattomin vertailukokein (Jadad & Enkin, 2007; Hróbjartsson ym., 2014). Ei kuitenkaan ole syytä päätellä, että tutkijat ja tutkimusten rahoittajat muokkaisivat tahallaan tutkimustuloksia mieleisekseen. Vaikka näinkin voidaan joskus tehdä, todennäköisesti yleisemmin tutkimusten harha on tahatonta, ja se jää usein jopa tutkijoilta huomaamatta (Jadad & Enkin, 2007).

Satunnaistetun vertailukokeen suurin vahvuus on tutkittavien tutkimusryhmiin valikoitumisesta aiheutuvan harhan hallitseminen (Hinkka-Yli-Salomäki, 2021). Hallitsemisella tarkoitetaan tässä yhteydessä eri harhalähteiden minimoimista tai jopa poistamista. Satunnaistamisella pyritään siihen, että tutkittavat ryhmät ovat mahdollisimman samankaltaisia tutkimuksen alussa. Näin tutkijan on lopuksi mahdollista eristää ja määrittää tutkittavan intervention vaikutus vertaamalla koeryhmää vertailuryhmään.

*”Satunnaistetun vertailukokeen suurin vahvuus on tutkimusryhmiin valikoitumisesta aiheutuvan harhan hallitseminen.”*

Millään muulla koeasetelmalla emme pysty tasapainottamaan tuntemattomia ennustavia tekijöitä tutkimuksen lähtötilanteessa. Tutkimukseen osallistuvien satunnainen kohdentaminen eri ryhmiin ei kuitenkaan suojaa satunnaistettuja vertailukokeita muun tyyppisiltä harhan lähteiltä. Ne voivat liittyä tutkimuksen suunnitteluun, toteutukseen tai raportointiin tai tiedon levittämiseen.

## MIKSI HARHALLA ON MERKITYSTÄ?

Mielenterveyden tukemiseen tarkoitettujen psykososiaalisten interventioiden todellisia vaikutuksia ei tunneta, kuten ei muidenkaan interventioiden. Tutkimuksessa harhaa koitetaan hallita, havaita, ennakoita ja määrittää, jotta tutkimustulokset olisivat mahdollisimman lähellä todellisuutta ja johtopäätökset voitaisiin näin ollen yleistää koskemaan laajaa kohdeväestöä.

On kuitenkin mahdotonta tietää varmasti, ovatko tietyt tutkimuksen tulokset harhaisia. Tämä johtuu yksinkertaisesti siitä, että on mahdotonta selvittää, eroavatko tutkimustulokset järjestelmällisesti siitä, mitä olisivat todelliset tulokset – joista ei voida saada tarkkaa tietoa. Tiedämme kuitenkin, että tietyt valinnat aiheuttavat harhaa toisia valintoja todennäköisemmin, ja näitä tutkimuksen eri osa-alueita on mahdollista arvioida (Sterne ym., 2019). Voimme tarkastella muun muassa:

- tutkittavien määrää ja valintaa
- vertailuryhmälle annettujen intervention valintaa
- vaikuttavuuden mittaamiseen käytettyjä mittareita
- satunnaistamisen toteuttamista
- sokkoutuksen tasoa
- käytettyjä tilastollisia menetelmiä.

Vertailemalla tutkimuksesta kieltäytyneiden ja tutkimuksen keskeyttäneiden määriä koe- ja vertailuryhmissä voimme lisäksi arvioida tutkimustulosten yleistettävyyttä ja soveltuvuutta kohde-ryhmään ja mahdollisesti myös laajempaan joukkoon. Satunnaistetuissa vertailukokeissa yleiset harhan lähteet voidaan jakaa viiteen luokkaan: 1) valinta-, 2) toiminta-, 3) havaitsemis-, 4) hiertymis- ja 5) raportointiharha (taulukko 1). Näiden ulkopuolelle jäävät harhalähteet yhdistetään usein omaan luokkaansa muut harhat.

### Valintaharha

Todellisessa satunnaistamisessa kaikilla tutkimukseen osallistuvilla on yhtäläinen mahdollisuus tulla valituksi mihin tahansa tutkimusryhmään. Toisin sanoen, jos tutkimuksessa on kaksi ryhmää, koe- ja vertailuryhmä, jokaisella osallistujalla on 50 prosentin todennäköisyys tulla satunnaistetuksi koeryhmään (ja vastaavasti 50 prosentin todennäköisyys tulla satunnaistetuksi vertailuryhmään). Usein tämä kuitenkin toteutuu vajavaisesti

Valintaharha (*selection bias*) voi syntyä tutkittavien rekrytoitavasta. Esimerkiksi, jos tutkija tietää tai arvaa satunnaistamisen

puutteellisen toteutuksen vuoksi seuraavaksi satunnaistettavan henkilön ryhmän, hän voi jättää osallistujaksi täysin kelvollisen henkilön tutkimuksen ulkopuolelle. Erityisesti tämä koskee avoimia (*open*) eli sokkouttamattomia tutkimuksia, kuten eri terapiamuotojen tutkimuksia, sekä tilanteita, joissa tutkija tietää satunnaistamisen sekvenssin eli lohkokoon. Lohkokoko tarkoittaa lukumäärää, jonka jälkeen koe- ja vertailuryhmän koot ovat samat. Jos esimerkiksi lohkokoko on neljä, tutkimuksen koe- ja vertailuryhmien koot ovat samat jokaisen neljän rekrytoituvan tutkittavan jälkeen. Toisin sanoen ensimmäisten neljän rekrytoituvan jälkeen koe- ja vertailuryhmissä on molemmissa kaksi tutkittavaa satunnaistettuna (kuvio 1).

*”On mahdotonta selvittää, eroavatko tutkimustulokset järjestelmällisesti siitä, mitä olisivat todelliset tulokset.”*

Valintaharha voi esiintyä myös vaiheessa, jossa jo tutkimukseen mukaan otettuja ollaan allokoimassa koe- ja vertailuryhmiin, jos tutkija tietää satunnaistamisessa käytetyn lohkokoon. Tällöin on mahdollista, että tutkija vaihtaa tutkimushenkilöiden satunnaistamisen järjestystä yrittäen saada esimerkiksi vaikeammin oireilevan henkilön vertailuryhmään, jotta koeryhmässä saataisiin paremmat tulokset lievempien tapausten ansiosta.

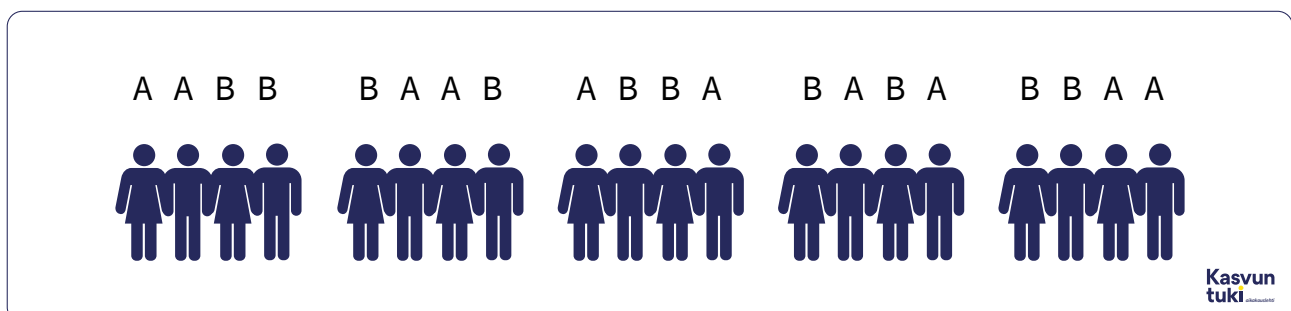
Schulz ja hänen tutkimusryhmänsä osoittivat jo vuonna 1995, että tutkimuksissa, joissa tutkijat tietävät satunnaistamiskoodin – eli heillä on tieto siitä, mitä interventiota kukin tutkittava saa – tai satunnaistamisen lohkokoon, tutkittavalle interventiolle saadaan keskimäärin 40 prosenttia suurempi vaikuttavuus kuin tutkimuksissa, joissa edellä mainitut seikat eivät ole tutkijoiden tiedossa (Schulz ym., 1995).

Satunnaistamiskoodin salaaminen on yksinkertaista, joten on erikoista, että sitä ei tutkimuksissa aina tehdä. Koodi voidaan ottaa käyttöön osana mitä tahansa koetta. Jos artikkelissa mainitaan, että satunnaistamiskoodi ja lohkokoko salattiin niiltä tutkimusta tekeviltä henkilöiltä, jotka olivat suorassa yhteydessä tutkittaviin, lukija voi vakuuttua, että valintaharha on mahdollisimman pieni. Läheskään aina tätä ei mainita, jolloin lukijan on mahdotonta tietää, toteutettiinko salaaminen vai ei.

### KUVIO 1.

#### Esimerkki satunnaistetusta lohkokokeesta, jossa lohkokoko on neljä

A ja B ovat koe- ja vertailuryhmille annettavat interventiot.



Satunnaistetun vertailukokeen raportoinnin CONSORT 2010 -tarkistuslistan mukaan satunnaistamisen yksityiskohtainen kuvaus, joka sisältää muun muassa satunnaistamislistan tekemiseen käytetyn menetelmäkuvauksen, on tärkeä osa tutkimuksen raportointia (Hinkka-Yli-Salomäki, 2021). Esimerkiksi kuvion 1 kaltaisessa tilanteessa satunnaistamislistassa on kaksi saraketta: ensimmäisessä sarakkeessa on tutkittavan tutkimusnumero (1–20) ja toisessa sarakkeessa tieto tutkittavan ryhmästä, eli joko koeryhmä (= A) tai vertailuryhmä (= B). Tämän osion puuttuminen tai puutteellinen kuvaus artikkelissa kertoo lukijalle tutkimuksen mahdollisesta valintaharhasta.

### Toimintaharha

Toimintaharhaa (*performance bias*) saattaa esiintyä, jos tutkija tai tutkittava tietää, mihin tutkimusryhmään tutkittava on satunnaistettu. Toimintaharhalle alttein satunnaistetun vertailukokeen tyyppi on täysin sokkouttamaton tutkimus, jossa sekä tutkija että tutkittava tietävät, mitä interventiota tutkittavalle annetaan.

## ”Intervention vaikutuksen raportoitu suuruus oli riippuvainen sokkoutuksen tasosta.”

Yksöissokkokeessa (*single-blind trial*) tutkittava ei tiedä mitä interventiota saa, mutta interventiota antava henkilö sen sijaan on tietoinen siitä, mihin ryhmään tutkittava on satunnaistettu. Kaksoissokkokeessa (*double-blind trial*) intervention toteuttaja ja tutkittava eivät tiedä, mitä interventiota tutkittava saa. Kaikkein tehokkaimmin toimintaharha minimoidaan kolmoissokkokeessa (*triple-blind trial*). Siinä satunnaistamiskoodia ei tiedä tutkittava, interventiota antava eikä tutkimusaineistoa analysoiva henkilö. Käytännössä tämä toteutetaan yleensä niin, että tilastotieteilijä analysoi päävasteen sokkona, jonka jälkeen satunnaistamiskoodi avataan myös hänelle (<http://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials>).

Kaksois- ja kolmoissokkokeissa toimintaharhaa ei ole tai se on hyvin pieni. Psykososiaalisen menetelmän vaikuttavuustutkimus on kuitenkin usein mahdotonta toteuttaa kaksois- tai kolmoissokkokeena. Esimerkiksi verrattaessa internetin välityksellä annettavaa terapiaa kasvokkain tapahtuvaan ryhmäterapiaan molemmat osapuolet, sekä osallistuja että terapeutti, tietävät, kummasta interventiosta on kyse.

Harha saattaa suurentaa intervention tulkittua vaikutusta, erityisesti jos tutkimuksen vasteet ovat subjektiivisesti arvioitavia eivätkä esimerkiksi mittalaitteella mitattavia arvoja (Savović ym., 2012). Eräässä systemaattisessa katsauksessa käytiin läpi 33 meta-analyysia, joissa oli mukana 250 satunnaistetun vertailukokeen julkaisuita. Tutkijat havaitsivat, että intervention vaikutuksen raportoitu suuruus oli riippuvainen sokkoutuksen tasosta ja riippuvuus oli tilastollisesti merkitsevä. Julkaisut, joissa ei kerrottu tutkimuksessa käytetyn kaksoissokkoutusta, raportoivat intervention vaikutukset keskimäärin 17 prosenttia suuremmiksi verrattuna julkaisuihin, joissa kerrottiin kaksoissokkoutusta käytetyn (Schulz ym., 1995).

Jos tutkimusta ei ole mahdollista sokkouttaa, tutkija voi käyttää muita keinoja harhan pienentämiseksi. Mikäli tutkimuksessa käytetään subjektiivisesti arvioitavia vasteita, toimintaharhaa voidaan pienentää sokkouttamalla arvioija. Esimerkiksi kognitiivisen käyt-

täytymisterapian (CBT) vaikutusta kartoittaneessa tutkimuksessa tutkittavan antamien vastausten arvioijana ei ollut interventiota antava tutkija, vaan toinen henkilö. Toisessa esimerkissä kaksi tutkimusryhmän ulkopuolista tutkijaa analysoivat ja pisteyttivät tutkijan videoimat tallenteet äidin ja vastasyntyneen vuorovaikutustilanteista.

Miten usein tutkimuksissa sitten mainitaan sokkoutuksesta tai siitä, ettei sitä käytetty? Tutkiessaan sokkoutuksen raportoinnin laatua Montori ym. (2002) arvioivat viiden arvostetun julkaisusarjan tulosjulkaisuja satunnaistetuista vertailukokeista. Tutkimusryhmä selvitti, oliko julkaisuissa mainintaa sokkoutuksesta ja siitä, ketkä olivat sokkona ja kuinka pitkän aikaa. Mahdollisia sokkoutettavia ryhmiä olivat tutkittavat, terveydenhuollon ammattihenkilöt, aineiston kerääjät, vasteiden arvioijat, tilastotieteilijät ja julkaisun kirjoittajat. Jonkin edellä mainitun ryhmän sokkouttamisesta mainittiin ainoastaan 25 prosentissa tutkituista satunnaistetuista vertailukokeista.

On tärkeää ymmärtää valinta- ja toimintaharhan välinen ero, eli mikä johtuu siitä, että satunnaistamiskoodia ei salata tutkimuksessa osallisina olevilta henkilöiltä, ja mikä taas sokkoutuksen puutteesta. Satunnaistamiskoodin salaaminen on aina mahdollista, sokkouttaminen ei.

### Havaitsemisharha

Havaitsemisharhalla (*detection bias*) tarkoitetaan järjestelmällisiä eroja tutkimusryhmien välillä siinä, miten vasteet mitataan tai arvioidaan. Sokkouttamattomissa tutkimuksissa tätä harhaa voidaan pienentää samalla tavalla kuin toimintaharhaa, eli käyttämällä tutkittavan raportoimien vastausten arvioijana toista henkilöä kuin interventiota antanutta tutkijaa. Arvioitsijoiden sokkouttaminen on erityisen tärkeää subjektiivisten tuntemusten, kuten tutkittavan ahdistuksen määrän, arvioinnissa (Philips ym., 2022).

## ”Arvioitsijoiden sokkouttaminen on erityisen tärkeää subjektiivisten tuntemusten arvioinnissa.”

Satunnaistetun vertailukokeen raportoinnin CONSORT 2010 -tarkistuslistan mukaan sokkoutuksen yksityiskohtainen kuvaus on osa laadukasta tutkimuksen raportointia (Hinkka-Yli-Salomäki, 2021). Kuvauksen tulee sisältää tiedot henkilöistä, jotka olivat tutkimuksen aikana sokkona, sekä siitä, kuinka sokkoutuksen säilyminen taattiin. Tiedon puuttuminen tai sen puutteellinen kuvaus kertoo lukijalle tutkimuksen mahdollisesta valinta- tai havaitsemisharhasta.

### Hiertymisharha

On aina mahdollista, että osa satunnaistettuun vertailukokeeseen rekrytoituista jää pois tutkimuksesta ennen sen päättymistä. Tavallisimpia tutkimuksesta poisjäämisen syitä ovat keskeyttäminen ja se, että tutkittavaan ei saada enää yhteyttä. Tutkittava voi jäädä pois esim. motivaation puutteen, avioeron, sairastumisen tai lähiomaisen sairastumisen tai kuoleman vuoksi. Joskus taas osallistuminen keskeytyy esimerkiksi tutkijan tekemään lastensuojeluilmoitukseen. Poisjääntejä voi tapahtua järjestelmällisesti tietyssä tutkittavien osajoukossa. Jos koe- ja vertailuryhmien keskeyttäneiden suhteellinen määrä on hyvin erilainen, hiertymisharhalla (*attrition bias*) on vaikutusta tutkimuksen tuloksiin ja raportointiin (Philips ym., 2021).

Satunnaistetun vertailukokeen raportoinnin CONSORT 2010 -tarkistuslistan mukaan jokaisen laadukkaan julkaisun pitäisi kuvata tutkimuksen osallistujavirta, eli kuinka monta tutkittavaa satunnaistettiin ja mihin ryhmiin sekä analyyseihin osallistuneiden määrä päävasteen osalta. Lisäksi pitäisi kuvata tutkittavien keskeyttämisten ja poissulkemisten määrät ja syyt kussakin ryhmässä (Hinkka-Yli-Salomäki, 2021). Tavallisesti luvut esitetään julkaisuissa vuokaavioina. Tietojen puuttuminen tai puutteellinen kuvaus kertoo lukijalle tutkimuksen mahdollisesta hiertymisharhasta.

### Raportointiharha

Raportointiharhalla (*reporting bias*) tarkoitetaan järjestelmällistä valintaa, mitä raportoidaan ja mitä jätetään raportoimatta. Satunnaistetusta vertailukokeesta tehtyä julkaisua lukiessaan jokainen voi arvioida mahdollista raportointiharhaa. Valikoiva raportointi on raportointiharhan yleisin muoto. Tämä johtuu siitä, että houkutus julkaista vain tilastollisesti merkitseviä tuloksia voi olla tutkijalle suuri. Lukija voi havaita valikoinnin vertaamalla julkaisua tutkimusrekisteriin ilmoitettuihin tietoihin vasteista. Toisin sanoen, lukija voi selvittää, miten yhtenevät tutkimusrekisterin tiedot päävasteista ja toissijaisista vasteista ovat julkaisun tietojen kanssa. Tällä hetkellä käytetyin kaikille avoin ei-kaupallinen tutkimusrekisteri on ClinicalTrials.gov (<https://clinicaltrials.gov/>).

CONSORT 2010 -tarkistuslistan mukaan kaikissa satunnaistetun vertailukokeen julkaisuissa tulisi mainita tutkimusrekisterin nimi ja rekisteröintinumero sekä tieto siitä, missä tutkimussuunnitelma on kokonaisuudessaan luettavissa, mikäli se on saatavilla (Hinkka-Yli-Salomäki, 2021). Tietojen puuttuminen kertoo lukijalle tutkimuksen mahdollisesta raportointiharhasta. Harha on vähäisintä satunnaistetuissa vertailukokeissa, joissa tutkittavien määrä on suuri, ja tutkimuksissa, joissa intervention vaikutus tutkittaviin vasteisiin on suuri verrattuna vertailuintervention vaikutuksiin.

Tulosten raportoimatta jättäminen, erityisesti päävasteiden osalta, on puutteellisesti tunnistettu ongelma. Ongelman vaikutus kertaantuu systemaattisissa katsauksissa: Kirkhamin ym. (2009) mukaan yli 55 prosentissa julkaistuista systemaattisista katsauksista oli puutteita päävasteen raportoinnissa tai päävastetta ei ollut raportoitu lainkaan.

### Muut harhat

Edellä mainittujen harhojen ulkopuolelle jää lukuisia tutkimuksen eri vaiheisiin liittyviä harhoja, joista tässä mainitaan muutama.

Esiinikaivuharhalla (*ascertainment bias*) tarkoitetaan otantaharhaa, joka liittyy aineiston hankintaan jonkin erityispiirteen tai ilmiönsuunnan perusteella (Mansournia ym. 2018). Toisin sanoen, tutkimukseen rekrytoidaan lähdeaineistoista joillakin erityispiirteillä valikoitunut otos huomioiden tutkimushenkilöitä koskevat sisäänotto- ja poissulkukriteerit. Esiinikaivuharhaa voi esiintyä myös rekrytoinnin jälkeisissä tutkimuksen vaiheissa ja sitä voi aiheuttaa mikä tahansa tutkimusprosessiin osallinen ryhmä.

Harhaa voi tuottaa:

- 1) interventiota antava henkilö: esimerkiksi tutkimushoitaja, perhevalmentaja tai terapeutti haluaa nähdä uudella interventiolla todellisuutta suuremman vaikutuksen
- 2) interventiota saava tutkittava: esimerkiksi yläkoululaisen (tai vastaavasti hänen huoltajansa) tietoinen oireiden aliarviointi tutkimukseen päästyä ("plasebovaikutus")
- 3) vasteita arvioiva tutkija: esimerkiksi intervention vaikutuksen ylikorostaminen arvioitaessa videotallennetta lapsen ja äidin vuorovaikutustilanteesta
- 4) tutkimusaineistoa analysoiva tilastotieteilijä: esimerkiksi tilastollisen mallin tai menetelmän valinta puuttuvan tiedon käsittelemiseksi siten, että se ylikorostaa intervention vaikutusta
- 5) tutkimusraportin kirjoittaja tutkimuskuvauksessaan.

## TAULUKKO 1.

### Yhteenveto tutkimusharhoista

Harha	Yhteenveto	Esimerkki
<b>Valintaharha</b>	Harha, joka aiheutuu menetelmästä, jolla tutkittavat kohdennetaan koet- tai vertailuryhmään.	Matalan kynnyksen hoito-ohjelman tutkimuksessa tutkija saattaa satunnaistaa vertailuryhmään henkilön, jolla on vaikeampia ja pitkäkestoisimpia masennusoireita, mikäli tutkija tietää tai voi arvata lohkokoon. Tällöin koeryhmän intervention vaikutus ylikorostuu, jos se vaikuttaa tehokkaammin lievemmin oirehtiviin tutkittaviin.
<b>Toimintaharha</b>	Harha, jossa tutkijat tai tutkittavat ovat tietoisia tutkittaville annettavista hoidoista, ja käyttäytyvät tämän seurauksena eri lailla.	Tutkittava tietää olevansa vertailuryhmässä, joka saa vain tietoa masennuksesta. Tämä voi vaikuttaa masennuskyselyitä täytettäessä niin, että oireita liioitellaan.
<b>Havaitsemisharha</b>	Harha, joka johtuu siitä, että vasteiden arvioijat ovat tietoisia annettavista hoidoista.	Tutkija tietää tutkittavien saavan interventiota ja arvioi tietoisesti heidän oireensa alakanttiin, mikä aiheuttaa intervention vaikutusten ylikorostumisen.
<b>Hiertymisharha</b>	Harha, joka aiheutuu siitä, että jokin tekijä aiheuttaa ei-satunnaista tutkimuksen keskeyttämistä.	Ero interventio- ja vertailuryhmien joissakin perustiedoissa (esimerkiksi sukupuoli, perherakenne, sosioekonominen asema) saattaa vaikuttaa siihen, että esimerkiksi pitkäkestoisen terapian ryhmässä keskeytetään odotettua useammin kuin lyhytkestoisen lähterapian ryhmässä.
<b>Raportointiharha</b>	Harha, joka johtuu osan tuloksista, tavallisesti tilastollisesti ei-merkitsevien tulosten, raportoimatta jättämisestä.	Clinicaltrials.gov -tutkimusrekisterin mukaan tutkimuksen päävaste on yleistä ahdistuneisuutta mittaava GAD-7. Tutkimuksesta julkaistu artikkeli ei kuitenkaan sisällä päävasteena yleistä ahdistuneisuutta, vaan keskittyy kahteen erityiseen ahdistuksen muotoon eli sosiaaliseen fobiaan ja eroahdistukseen.

TAULUKKO 2.

Satunnaistettujen vertailukokeiden harhan arvioiminen Cochranen harhan riski -työkalulla

Harha ja siihen liittyvät kysymykset	Vastausvaihtoehdot		
	Pieni riski	Suuri riski	Muu
<b>1. Satunnaistamisprosessiin liittyvä harha</b>			
1.1 Oliko satunnaistamisen järjestys (sekvenssi) satunnaistettu?	K	E	-
1.2 Salattiinko satunnaistamisen järjestys (sekvenssi), kunnes tutkittava oli otettu tutkimukseen mukaan ja satunnaistaminen tehty?	K	E	-
1.3 Indikoivatko tutkimusryhmien perustason erot ongelmaa satunnaistamisprosessissa?	K	E	-
Harhan riskin arvio (pieni/suuri/jonkin verran)			
Valinnainen: Mikä on satunnaistamisprosessiin liittyvän harhan suunta?			
<b>2. Aiottujen interventioiden poikkeamisiin liittyvä harha</b>			
2.1 Olivatko tutkittavat tietoisia heille määrätystä interventiosta tutkimuksen aikana?	E	K	-
2.2 Olivatko interventiota antavat henkilöt tietoisia tutkittaville määrätystä interventiosta tutkimuksen aikana?	E	K	-
2.3 Jos ”K” kohdassa 2.1 tai 2.2, niin johtuiko poikkeama tutkimuksen kontekstista?	E	K	-
2.4 Jos ”K” kohdassa 2.1 tai 2.2, niin saattoivatko poikkeamat vaikuttaa tuloksiin?	E	K	-
2.5 Jos ”K” kohdassa 2.1 tai 2.2, niin olivatko poikkeamat tutkimusryhmien välillä tasapainossa?	K	E	-
2.6 Käytettiinkö sopivaa analyysia, kun estimoitiin hoitoaikkeen mukaisia vaikutuksia?	K	E	-
2.7 Jos ”K” kohdassa 2.6, niin vaikuttiko tutkimuksen tulokseen se, että kaikkia satunnaistamisjärjestyksen mukaisesti satunnaistettuja ei voitu analysoida?	E	K	-
Harhan riskin arvio (pieni/suuri/jonkin verran)			
Valinnainen: Mikä on aiottujen interventioiden poikkeamisiin liittyvän harhan suunta?			
<b>3. Puuttuvan tiedon harha</b>			
3.1 Olivatko tulokset olemassa kaikilta tai lähes kaikilta, jotka satunnaistettiin?	K	E	-
3.2 Jos ”E” kohdassa 3.1, onko todisteita, ettei analyysin tulos ole puuttuvista tiedoista johtuen harhainen?	K	E	-
3.3 Jos ”E” kohdassa 3.2, voisiko tiedon puuttuminen olla riippuvainen vasteen todellisesta arvosta?	E	K	-
3.4 Jos ”K” kohdassa 3.3, onko todennäköistä, että tiedon puuttuminen on riippuvainen vasteen todellisesta arvosta?	E	K	-
Harhan riskin arvio (pieni/suuri/jonkin verran)			
Valinnainen: Mikä on puuttuvaan tietoon liittyvän harhan suunta?			
<b>4. Vasteen mittauksen harha</b>			
4.1 Oliko vasteen mittaamiseen käytetty menetelmä sopimaton?	E	K	-
4.2 Oliko mahdollista mitata vaste eri tavoin tutkimuksen eri ryhmissä?	E	K	-
4.3 Jos ”E” kohdassa 4.1 ja 4.2, olivatko mittajaajat tietoisia osallistujien saamista interventiosta?	E	K	-
4.4 Jos ”K” kohdassa 4.3, oliko mahdollista, että vasteen arviointiin vaikutti tieto saadusta interventiosta?	E	K	-
4.5 Jos ”K” kohdassa 4.4, oliko todennäköistä, että vasteen arviointiin vaikutti tieto saadusta interventiosta?	E	K	-
Harhan riskin arvio (pieni/suuri/jonkin verran)			
Valinnainen: Mikä on vasteen mittaukseen liittyvän harhan suunta?			
<b>5. Raportoidun tuloksen valintaan liittyvä harha</b>			
5.1 Oliko aineisto analysoitu ennen satunnaistamiskoodin avaamista hyväksytyin analyysisuunnitelman mukaisesti?	K	N	-
Onko raportoitu tulos mahdollisesti valittu sellaisten tulosten perusteella, jotka			
5.2 on saatu useista vasteen mittauksista (esimerkiksi monikysymysmittarit, useat aikapisteet)?	E	K	-
5.3 on saatu useista saman vasteen eri analyysitavoista?	E	K	-
Harhan riskin arvio (pieni/suuri/jonkin verran)			
Valinnainen: Mikä on raportoidun tuloksen valintaan liittyvän harhan suunta?			
<b>Harhan kokonaisarvio</b>			
Harhan riskin arvio (pieni/suuri/jonkin verran)			
Valinnainen: Mikä on tämän vasteen arvioidun kokonaisharhan suunta?			

K = Kyllä tai todennäköisesti kyllä. E = Ei tai todennäköisesti ei. - = Ei tietoa tai ei sovellettavissa. Lähde: Cochrane Risk of Bias Tool (Sterne ym., 2019).

Paras tapa esiinkaivuharhan minimoimiseksi on pitää tutkimukseen osallistuvat henkilöt tietämättöminä annetusta interventiosta mahdollisimman kauan. Satunnaistaminen ja sokkoutus ovat siis tämänkin harhan minimoimisessa avainasemassa.

Muodikkuisharhaa (*hot stuff bias*) voi esiintyä tilanteissa, jossa tutkimusaihe on noussut tutkijoiden keskuudessa muodikkaaksi. Tällöin tutkijat voivat olla vähemmän kriittisiä ja artikkelit voidaan saada julkaistua nopeammin ja helpommin. Esimerkkinä tästä ovat koronapandemian alussa tehdyt lukuisat tapausartikkelit ja artikkelit, joissa mainitaan tekoäly tai koneoppiminen sovellusalasta riippumatta. Julkisuudella pystytään muokkaamaan nopeasti ja tehokkaasti muoti-ilmiöiksi nousevia aiheita.

## ”Kirjoittajan ja lukijan kannattaa käyttää raportoinnin CONSORT 2010 -tarkistuslistaa ja Cochranen harhan riski -työkalua.”

Rahoittamisharhaa (*funding bias*) voi esiintyä tilanteissa, joissa tutkija pyrkii analysoimaan ja raportoimaan tutkimuksensa puolueellisesti tukien rahoittajan tarpeita. Historiasta tunnetaan tupakka- ja lääketieteellisuuden rahoittamia hyvinkin harhaisia tutkimuksia. Satunnaistetun vertailukokeen raportoinnin CONSORT 2010 -tarkistuslistan mukaan rahoituslähteet ja muu tuki sekä rahoittajien rooli tulisi mainita jokaisessa satunnaistetun vertailukokeen julkaisussa (Hinkka-Yli-Salomäki, 2021). Lisäksi tavallisesti artikkelin julkaisemisen edellytyksenä on, että kirjoittajat raportoivat taloudelliset ja muut sidonnaisuudet. Tietojen puuttuminen kertoo lukijalle tutkimuksen mahdollisesta rahoittamisharhasta.

Otoskokoharhalla (*wrong sample size bias*) viitataan sattumalöydöksiin otoskooltaan pienissä tutkimuksissa ja toisaalta suurten tutkimusten tilastollisesti merkitseviin tuloksiin, joilla ei kuitenkaan ole kliinistä käytännön merkitystä. CONSORT 2010 -tarkistuslistan mukaan otoskoon määrittely ja päävasteen ja mahdollisten toissijaisten vasteiden efektikokoko (ts. vaikutuksen koko) tulisi mainita satunnaistetun vertailukokeen julkaisussa (Hinkka-Yli-Salomäki, 2021).

## HARHAN ARVIOIMINEN COCHRANEN HARHAN RISKI -TYÖKALULLA

Satunnaistettujen vertailukokeiden harhojen lähteitä ja suuruutta arvioidaan systemaattisissa katsauksissa yleisesti Cochranen harhan riski -työkalulla. Ensimmäinen versio työkalusta on vuodelta 2008. Tätä nykyä käytössä on päivitetty versio vuodelta 2019 (Cochrane Risk of Bias Tool; Sterne ym., 2019). Työkalu sisältää viisi osa-aluetta, joilla mahdollinen harha voi vaikuttaa tutkimuksen tuloksiin (taulukko 2). Harhan riskiä kullakin osa-alueella arvioidaan asteikolla:

- pieni tai todennäköisesti pieni riski
- suuri tai todennäköisesti suuri riski
- ei tietoa tai ei sovellettavissa

Harhan riski arvioitavassa tutkimuksessa saadaan arvioimalla kutakin viittä osa-aluetta. Näin tutkimukset saadaan luokiteltua harhan riskin mukaan. Jos riski arvioidaan kaikilla osa-alueilla pieneksi, kokonaisriski on pieni. Jos taas riski arvioidaan suureksi edes yhdellä osa-alueella, kokonaisriski on suuri. Jos riski on suurentunut vähintään yhdellä osa-alueella, mutta ei yllä suurimmalle tasolle yhdelläkään osa-alueella, kokonaisriskiksi saadaan ”jonkin verran”.

## ”Kliinisten tutkijoiden ja epidemiologien käyttämä terminologia on erilaista.”

Satunnaistetusta vertailukokeesta tehtyä artikkelia lukiessa kannattaa arvioida edellä esitettyjen harhojen riskejä. On kuitenkin hyvä muistaa, ettei täysin harhatonta tutkimusta ole olemassa. Harhalla on suora vaikutus siihen, miten luotettavina tuloksia voidaan pitää. Lukija arvioi tutkimuksen harhaa yleensä vain artikkelin perusteella, joten huonolaatuinen raportointi voi herättää epäilykset, vaikka itse tutkimus olisi toteutettu lähes harhattomasti (<https://training.cochrane.org/handbook/current>). Tämän vuoksi artikkelien kirjoittajien kannattaa käyttää käytettyä sekä satunnaistetun vertailukokeen raportoinnin CONSORT 2010 -tarkistuslistaa että Cochranen harhan riski -työkalua.

### TAULUKKO 3.

#### Kliinisten kokeiden harhatyyppien ja epidemiologisten termien vastaavuus

Cochranen <sup>1</sup> harhaluokitus	Epidemiologinen termi	Harha analyysissa	
		Hoitoaikeeseen mukainen	Tutkimussuunnitelman mukainen
Valintaharha	Sekoittuminen tai valintaharha	Kyllä	Kyllä
Toimintaharha	Sekoittuminen	Ei	Kyllä
Havaitsemisharha	Mittausharha	Kyllä	Kyllä
Hiertymisharha	Valintaharha	Kyllä	Kyllä
Raportointiharha	Raportointiharha	Kyllä	Kyllä

<sup>1</sup> <https://training.cochrane.org/handbook/current>

## Satunnaistetun vertailukokeen ja epidemiologisen havaintotutkimuksen eroja

Satunnaistettuja vertailukokeita ja havaintotutkimuksia käytetään interventoiden kausaalisten vaikutusten selvittämiseen. Näiden kahden lähestymistavoissa on kuitenkin useita eroja. Kuten nimikin kertoo, yksi eroista on se, että satunnaistetussa vertailukokeessa käytetään valintaharhan minimoimiseksi satunnaistamista. Parhaimmillaan aineisto analysoidaan hoitoaieperiaatteen (ts. ITT-periaate, *intention-to-treat*) mukaisesti. Tällöin mukaan analyysiin otetaan kaikki satunnaistetut tutkittavat, myös ne, jotka keskeyttivät tutkimuksen tai syystä tai toisesta saivat eri intervention kuin heille oli lähtöryhmän mukaisesti allokoitu.

**”Mikään tilastollinen menetelmä ei pysty poistamaan analyysia edeltävissä vaiheissa syntyneitä harhaa.”**

Toinen satunnaistetun vertailukokeen ja havaintotutkimuksen ero on sokkouttaminen. Sokkouttaminen ja hoitoaieperiaatteen mukainen analysointi alkoivat levitä laaja-alaiseen käyttöön vasta 1990-luvun lopulla viranomaisohjeiden myötä (<http://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials>). Ohjeet koskevat uusien lääke- ja diagnostiikkalaitteiden maailmanlaajuisia rekisteröintejä, mutta niitä on alettu soveltaa enenevässä määrin myös psykososiaalisten menetelmien vaikuttavuutta selvittämissä satunnaistetuissa vertailukokeissa.

Kliinisten tutkijoiden ja epidemiologien käyttämä terminologia on erilaista, mikä voi mutkistaa keskustelua satunnaistettujen vertailukokeiden ja havaintotutkimusten välisistä eroista. Kliiniset tutkijat luokittelevat harhat usein viiteen luokkaan. Epidemiologit sen sijaan käyttävät käsitteitä sekoittuminen, valintaharha ja mitausharha. Raportointiharhan nimitys ja merkitys ovat samat sekä epidemiologisessa että kliinisessä tutkimuksessa (Mansournia ym., 2018). Termien vastaavuudet on kuvattu taulukossa 3.

Valintaharha voi vaikuttaa sekä hoitoaiekeen että tutkimussuunnitelman mukaisen (*per-protocol*, PP) vaikutuksen estimaattiin (taulukko 3). Tutkimussuunnitelman mukaisessa aineistossa sisäänotto- ja poissulkukriteerien täytyminen huomioidaan koko tutkimuksen ajan ja tutkittavat osallistuvat kaikkiin suunnitelman mukaisiin vaiheisiin. Kuviossa 2 on havainnollistettu, miten tutkimussuunnitelman mukainen aineisto sisältyy aina hoitoaieperiaatteen mukaiseen aineistoon. PP- ja ITT-periaatteiden välinen ero tutkimushenkilöiden lukumäärässä vaihtelee tutkimuksesta toiseen, mutta toisinaan se voi olla useita prosentteja.

Valintaharhaa ilmenee, jos koe- ja vertailuryhmien alkumittaukset eroavat merkittävästi toisistaan. Tällöin tutkittavien ennustekijät (*prognostic factors*) ovat vaikuttaneet tutkittavien allokointiin koe- ja vertailuryhmään (Mansournia ym., 2018).

Toimintaharhaa, eli epidemiologisin termein sekoittumista, ei ole hoitoaieperiaatteen mukaisessa analyysissa (taulukko 3). Havaitsemis-, hiertymis- ja raportointiharha sen sijaan ovat mahdollisia niin ITT- kuin PP-periaatteen mukaisissa analyyseissa. Yhteenvedon voidaan todeta, että satunnaistettu vertailukoe kannattaa analysoida ottamalla mukaan kaikki satunnaistetut tutkittavat ITT-periaatteen mukaisesti. Tämä minimoi toimintaharhan. Kuvio 3 havainnollistaa hoitoaiekeen ja tutkimussuunnitelman mukaista vasten arvioinnin eroa.

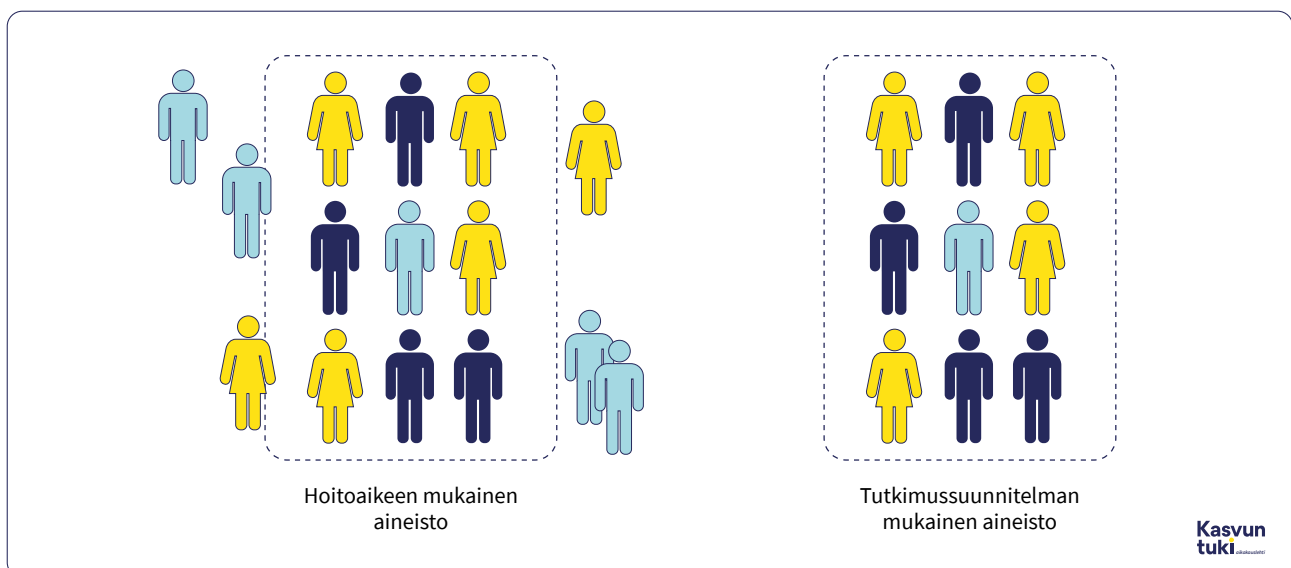
**”Sokkouttaminen ja satunnaistamisessa käytetyn lohkokoon salaaminen ovat yleisimpiä keinoja minimoida harhaa.”**

### LOPUKSI

On hyvä muistaa, että täysin harhatonta satunnaistettua vertailukoetta ei ole, eikä mikään tilastollinen menetelmä pysty poistamaan harhaa, jos sitä on esiintynyt analyysia edeltävissä vaiheissa. Satunnaistetun vertailukokeen suurin vahvuus on tutkittavien tutkimusryhmiin valikoitumisesta aiheutuvan harhan hallitseminen.

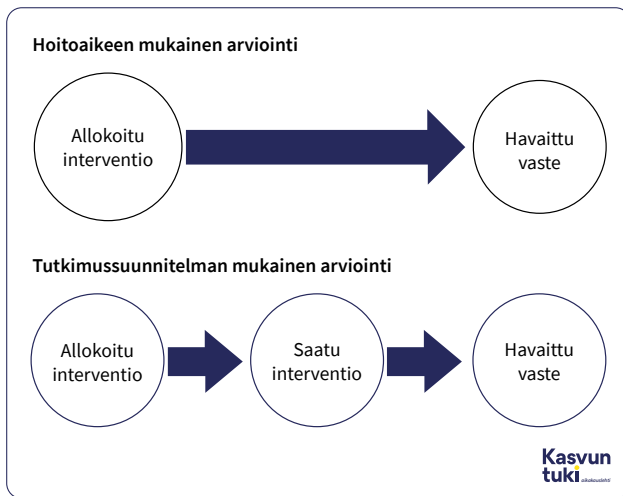
### KUVIO 2.

#### Hoitoaiekeen (Intention-to-treat) ja tutkimussuunnitelman mukaisen (per-protocol) aineiston havainnollistus



### KUVIO 3.

#### Hoitoaikeen (intention-to-treat) ja tutkimussuunnitelman mukaisen (per-protocol) vaikutuksen ero vasteen arvioinnissa



Sokkouttaminen ja satunnaistamisessa käytetyn lohkokoon salaaminen tutkimusta suorittavilta ja analysoivilta henkilöiltä sekä tutkitavilta ovat yleisimpiä keinoja minimoida harhaa. Mahdollisimman monen tahon sokkouttaminen pienentää harhaa, ja vähintään analysointivaiheen voi aina tehdä sokkona.

Sokkouttamattomien tutkimusten tulosten tulkinnassa on hyvä olla tarkkana. Niin avoimissa kuin sokkoutetuissakin vertailukokeissa objektiivisten ja luotettavien vasteiden ja mittareiden käyttö sekä vertailuryhmälle annetun intervention standardoiminen lisää tutkimuksen laatua. Standardoimisella tarkoitetaan tässä laadun varmistamista: esimerkiksi kaikkien interventiota antavien yhtäläistä kouluttamista samalla koulutusmateriaalilla ja intervention pitämistä täysin samanaikaisesti tutkimuksen alusta loppuun jokaisella tutkimushenkilöllä.

Lukija voi arvioida julkaistujen tulosten luotettavuutta, merkittävyyttä ja yleistettävyyttä satunnaistetun vertailukokeen raportoinnin CONSORT 2010 -tarkistuslistan (Hinkka-Yli-Salomäki, 2021) sekä Cochranen harhan riski -työkalun avulla (taulukko 1). Satunnaistetun vertailukokeen toteuttamisen vuokaavio, joka esittää osallistujavirran yksityiskohtaisesti, auttaa tulosten yleistettävyyden arvioinnissa (Hinkka-Yli-Salomäki, 2021). ●

### AVAINSANAT:

satunnaistettu vertailukoe, harha, raportointi, artikkeli, laatu

### SIDONNAISUUDET:

Sanna Hinkka-Yli-Salomäki: Ei sidonnaisuuksia.

### LÄHTEET:

- 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: Satunnaistetun vertailukokeen laadukkaan raportoinnin edellytykset. Kasvun tuki. Haettu 1.11.2022 osoitteesta <https://kasvuntuki.fi/aikakauslehti/julkaisut/kasvun-tuki-aikakauslehti-12022/hinkka-yli-salomaki/>
- Hróbjartsson, A., Emanuelsson, F., Skou Thomsen, A. S., Hilden, J. & Brorson, S. (2014). Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. *Int J of Epid*, 43(4), 1272–1283. <https://doi.org/10.1093/ije/dyu115>
- Jadad, A. R. & Enkin, M. W. (2007). *Randomized Controlled Trials: Questions, Answers, and Musings*. 2nd edition. Blackwell Publishing. <https://doi.org/10.1002/9780470691922>
- Karanicolas, P. J., Farrokhyar, F. & Bhandari, M. (2010). Practical tips for surgical research: blinding: who, what, when, why, how?. *Can J Surg*, 53(5), 345–348. <https://pubmed.ncbi.nlm.nih.gov/20858381/>
- Kirkham, J. J., Dwan, K. M., Altman, D. G., Gamble, C., Dodd, S., Smyth, R. & Williamson, P. R. (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*, 340. <https://doi.org/10.1136/bmj.c365>
- Mansournia, M. A., Higgins, J. P. T., Sterne, J. A. C. & Hernán, M. A. (2018). Biases in randomized trials: conversation between trialists and epidemiologists. *Epidemiology*, 28(1), 54–59. <https://doi.org/10.1097/EDE.0000000000000564>
- Montori, V. M., Bhandari, M., Devereaux, P. J., Manns, B. J., Ghali, W. A., & Guyatt, G. H. (2002). In the dark: the reporting of blinding status in randomized controlled trials. *J Clin Epidemiol*, 55(8), 787–790. [https://doi.org/10.1016/S0895-4356\(02\)00446-8](https://doi.org/10.1016/S0895-4356(02)00446-8)
- Phillips, M. R., Kaiser, P., Thabane, L., Bhandari, M. & Chaudhary, V. for the Retina Evidence Trials International Alliance (R.E.T.I.N.A.) Study Group. (2022). Risk of bias: why measure it, and how?. *Eye*, 36(2), 346–348. <https://doi.org/10.1038/s41433-021-01759-9>
- Savović, J., Jones, H. E., Altman, D. G., Harris, R. J., Jüni, P., Pildal, J., ... & Sterne, J. A. C. (2012). Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health technology assessment*, 16(35), 1–82. <https://doi.org/10.3310/hta16350>
- Schulz, K. F., Chalmers, I., Hayes, R. J. & Altman, D. G. (1995). Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of American Medical Association*, 273(5), 408–412. <https://doi.org/10.1001/jama.273.5.408>
- Sterne, J. A., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., ... & Higgins, J. P. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*, 366. <https://doi.org/10.1136/bmj.l4898>
- The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions*, Version 6.2. Chapter 8: Risk of Bias in Randomized Trials. Published 2021. <https://training.cochrane.org/handbook/current>