

tieteellisen tutkimuksen tikapuut



Ari Uusaro

Dosentti, MHS (Epid), apulaisylilääkäri
KYS, Ensihoito, päivystysalue ja tehohoito
ari.uusaro@ajkuh.fi



Tero Martikainen

LT, erikoistuva lääkäri
KYS, Ensihoito, päivystysalue ja tehohoito
tero.martikainen@ajkuh.fi

META-ANALYYSIEN ARVO JA MAHDOLLISET HEIKKOUDET TUTKIMUSTIEDON ARVIOINNISSA

Meta-analyysissä voidaan yhdistää useita samaa asiaa selvittäneitä pieniä aineistoja ja näin lisätä tutkimuksen voimaa interventio- ja kontrolliryhmän välisen mahdollisen eron löytämiseksi.

► Selvitettäessä hoitointerventio- ja kontrolliryhmän välisen mahdollisen eron löytämiseksi. Näin menetellen tutkimuksia yhdistettäessä tuloksen piste-estimaatin (= tuloksen) ympärillä oleva epävarmuusalue (=luottamusväli) pienenee.

Meta-analyysi eroaa systemaattisesta katsauksesta siinä, että meta-analyysissä tulokset esitetään aina kvantitatiivisesti analyysiin soveltuvia tilastollisia menetelmiä käyttäen. Meta-analyysien hyöty voi olla esimerkiksi se, että yhdistämällä useita pienempiä tutkimuksia meta-analyysi voi osoittaa intervention hyödyn ennen laajan ja kalliin yksittäisen satunnaistetun tutkimuksen tekemistä. Tässä tapauksessa potilaille hyödyllinen hoitomenetelmä voitaisiin ottaa käyttöön jo aikaisemmin. Tästä teoreettisesta hyödystä on esimerkki

tutkimuksen voimaa interventio- ja kontrolliryhmän välisen mahdollisen eron löytämiseksi. Näin menetellen tutkimuksia yhdistettäessä tuloksen piste-estimaatin (= tuloksen) ympärillä oleva epävarmuusalue (=luottamusväli) pienenee.

Meta-analyysi eroaa systemaattisesta katsauksesta siinä, että meta-analyysissä tulokset esitetään aina kvantitatiivisesti analyysiin soveltuvia tilastollisia menetelmiä käyttäen. Meta-analyysien hyöty voi olla esimerkiksi se, että yhdistämällä useita pienempiä tutkimuksia meta-analyysi voi osoittaa intervention hyödyn ennen laajan ja kalliin yksittäisen satunnaistetun tutkimuksen tekemistä. Tässä tapauksessa potilaille hyödyllinen hoitomenetelmä voitaisiin ottaa käyttöön jo aikaisemmin. Tästä teoreettisesta hyödystä on esimerkki

kuvassa 1. Tässä esimerkissä meta-analyysillä olisi todettu trombolyytihoitoon hyöty sydänlihaskeuhkoiden hoidossa jo useita vuosia ennen kuin asia varmistettiin satunnaistetuissa tutkimuksissa (1-3).

Meta-analyysien määrä on lisääntynyt 90-luvun loppua kohti ja lisääntyy edelleen (kuva 2) (4, 5). Meta-analyysija on kuitenkin myös kritisoitu voimakkaasti ja on viitattu jopa "tilastotieteelliseen alkemiaan" (6). Jossain määrin tämä kritiikki voi olla perusteltua, ja meta-analyysien laatu onkin usein todettu huonoksi (5-9). On myös todettu, että tietyn intervention vaikuttavuutta selvittävien meta-analyysien ja myöhemmin tätä samaa interventiota selvittävien laajojen satunnaistettujen tutkimusten tulokset eivät aina ole yhteneväiset. Kuvassa 3 on esimerkki

Kliinikolle onkin tärkeää pystyä erottamaan huonolaatuiset meta-analyysit laadukkaista.

sevän eron havaitsemiseen, vaikka ero olisi todellisuudessa olemassa. Meta-analyysissä on mahdollista yhdistää useita samaa asiaa selvittäneitä pieniä tutkimuksia ja siten lisätä

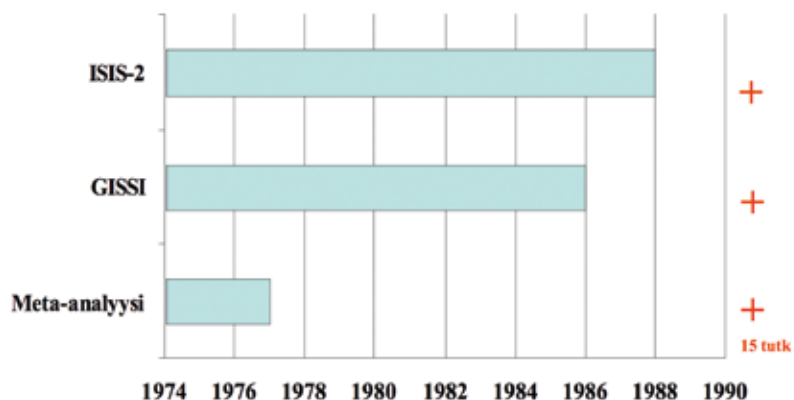
tutkimuksista, joissa selvitettiin magnesiumin iv-annon vaikuttavuutta uhkaavan sydäninfarktin hoidossa (10, 11). Aiemman meta-analyysin perusteella suositeltiin, että hoito olisi viipymättä otettava käyttöön, mutta myöhemmin tehdyssä laajassa tutkimuksessa hoito todettiin hyödyttömäksi. On osoitettu, että meta-analyysit eivät ennustaneet oikein laajan satunnaistetun tutkimuksen lopputulosta 35%:ssa tapauksista (12, 13). Kliinikolle onkin tärkeää pystyä erottamaan huonolaatuiset meta-analyysit laadukkaista.

Tämän artikkelin tarkoituksena on kuvata lyhyesti, kuinka meta-analyysi tehdään. Erityisenä tavoitteena on auttaa lukijaa huomaamaan, mitkä seikat meta-analyysissä saattavat vaarantaa tulosten validiteetin ja tutkimuksesta tehtävien johtopäätösten oikeellisuuden.

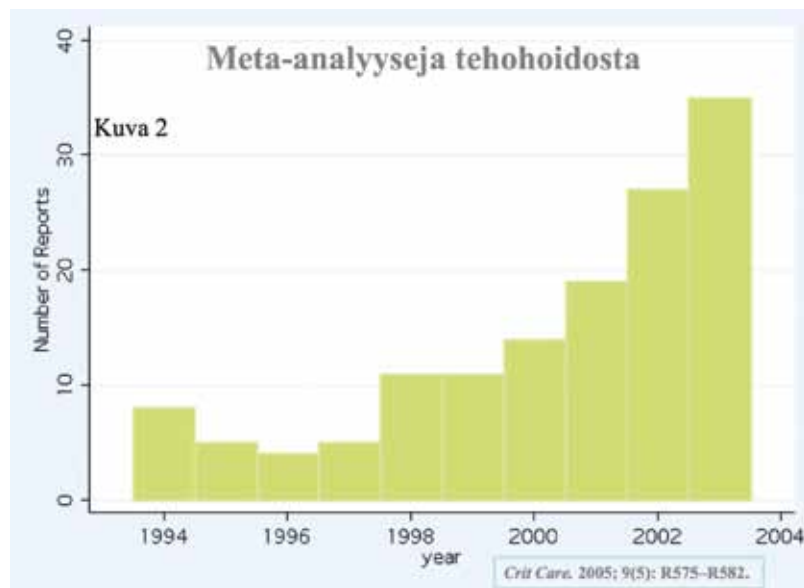
Tutkittavan aiheen valinta ja yksittäisten tutkimusten hakeminen

Ensimmäinen tehtävä meta-analyysissä on muotoilla kliinisesti relevantti ja mahdollisimman täsmällisesti kohdennettu kysymyksenasettelu. Interventio (esimerkiksi lääkehoito, hengityslaittehoito), päävaste- muuttuja (esimerkiksi kuolema, elämänlaatu, hoitopäivien määrä), potilasvalinta (esimerkiksi ARDS, pneumonia, sepsiksen vaikeusaste) ja kontrollipotilaat (esimerkiksi sillä hetkellä annettava rutiinihoito, "current practice") pitäisi pystyä määrittelemään mahdollisimman tarkasti. Mitä selkeämpi ja yksityiskohtaisempi kysymyksenasettelu on, sitä helpompi on myöhemmin pohtia esimerkiksi tulosten yleistettävyyttä ja niiden mahdollista vaikutuksesta hoitokäytäntöihin (14). Tämän jälkeen on määritettävä yksittäisten tutkimusten selkeät inklusio- ja eksklusiokriteerit ja tunnistettava meta-analyysiin mukaan otettavat artikkelit. Laadukkaassa meta-analyysissä myös raportoidaan, missä määrin eri tutkijat ovat olleet yksimielisiä yksittäisten artikkeleiden sisällyttämisestä meta-analyysiin.

Meta-analyysiin etsittävien tutkimusten hakumenetelmät tulee yksityiskohtaisesti kuvata ja esimerkiksi



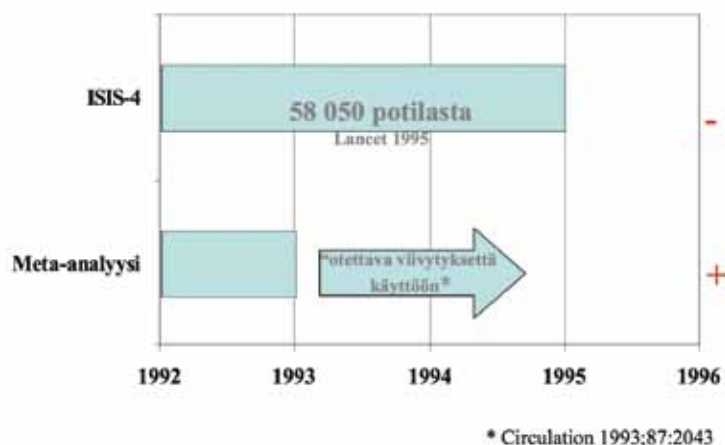
Kuva 1. Meta-analyysin avulla voidaan havaita kliinisesti merkittävä intervention hyöty ennen löydöstä vahvistavien laajojen satunnaistettujen tutkimusten julkaisemista. Tässä esimerkissä on tutkittu trombolyyysin tehoa uhkaavan sydäninfarktin hoidossa. ISIS-2 and GISSI ovat yksittäisiä satunnaistettuja tutkimuksia, jotka osoittivat intervention hyödyllisyyden. Meta-analyysillä tämä hyöty olisi havaittu useita vuosia ennen näiden yksittäisten tutkimusten julkaisemista (1-3). + = tutkimus osoitti tutkittavasta interventiosta olevan hyötyä (1, 2, 3).



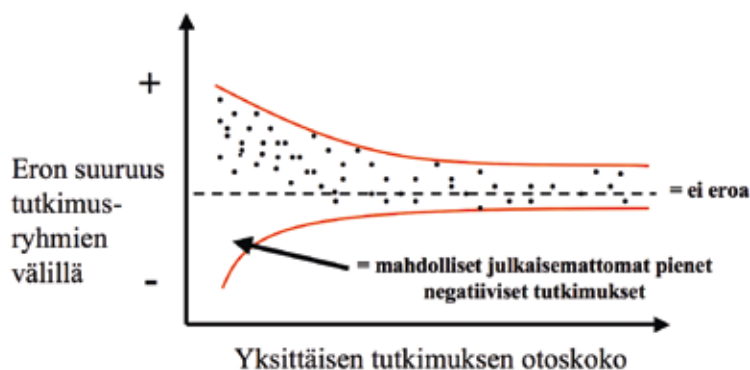
Kuva 2. Julkaistujen meta-analyysien määrä lisääntyy voimakkaasti. Esimerkki tehohoidosta (5).

kaikki huomioidut kirjallisuustietokannat täytyy raportoida (14, 15). Myös muut käytetyt menetelmät tulee kuvata: onko oltu yhteydessä esimerkiksi teollisuuteen tai yksittäisiin muihin tutkijoihin henkilökohtaisesti (14-16). Laadukkaat meta-analyysit

eivät rajoitu ainoastaan englanninkieliseen kirjallisuuteen (17, 18). On todettu, että tilastollisesti merkitsevät tulokset julkaistaan todennäköisemmin englannin kielellä kuin muulla kielellä (riskisuhte 3.75, 95% luottamusväli CI 1.25-11.3) (17). Näin >>



Kuva 3. Esimerkissä tutkittiin magnesiumin iv-annon hyötyä sydäninfarktipotilailla. Meta-analyysin jälkeen arvostetussa kardiologian alan lehdessä suositeltiin magnesiumin ottamista välittömästi kliniseen käyttöön. Meta-analyysissä saatiin kuitenkin väärä positiivinen tulos. Magnesiumin hyödyttömyys todettiin hyvin laajassa (58 050 potilasta) myöhemmin tehdyssä satunnaistetussa, kontrolloidussa tutkimuksessa. + = magnesiumista havaittiin olevan hyötyä, - = magnesiumista ei havaittu olevan hyötyä (10, 11).



Kuva 4. Esimerkki julkaisuharhasta (= publication bias). Tutkimukset julkaistaan todennäköisemmin, jos ne ovat tilastollisesti merkitseviä. Eryisesti näin on pienten tutkimusten kohdalla ja tällöin yksittäisessä tutkimuksessa ryhmien välisen eron tulee olla suuri. Pieniä negatiivisia tutkimuksia on vaikeampi saada julkaistuksi, eivätkä ne päädy meta-analyysiin. Jos meta-analyysissä on mukana useita pieniä positiivisia tilastollisesti merkitseviä tutkimuksia, niin tämän myötä intervention teho saatetaan yliarvioida. Yksittäinen piste = yksi satunnaistettu kontrolloitu tutkimus. + = suosii uutta interventiota, - = suosii kontrolliryhmää.

ollen muun kuin englanninkielisen kirjallisuuden poisjättäminen saattaa vääristää meta-analyysin tuloksia.

Julkaisuharha (= publication bias) on myös uhka meta-analyysin tulosten validiteetille (19-21). Julkaisuharha tarkoittaa sitä, että tilastollisesti merkitsevät tulokset julkaistaan todennäköisemmin kuin ei-merkitsevät tulokset. Jos ei-merkitsevät tutkimukset ovat sinänsä laadukkaita ja jos niitä ei saada mukaan meta-analyysiin, meta-analyysin validiteetti voi kärsiä. On osoitettu, että tilastollisesti merkitsevä tutkimustulos julkaistaan 2.3 (95% CI 1.25-4.28) kertaa todennäköisemmin kuin tilastollisesti ei-merkitsevä tulos (20). Sama havainto on tehty myös muissa selvityksissä ja lisäksi on todettu, että tilastollisesti merkitsevät tutkimukset julkaistaan nopeammin tutkimuksen loppumisen jälkeen verrattuna tilastollisesti ei-merkitseviin tutkimuksiin (20, 22). Myös otoskooltaan pieni tutkimus julkaistaan todennäköisemmin, jos tulos on tilastollisesti merkitsevä. Pienellä otoskoolta tehtyjen tutkimusten tulokset ovat tilastollisesti merkitseviä, jos tutkittavien ryhmien välinen ero tutkittavassa päävastemuuttujassa on suuri. Näin ollen jos meta-analyysissä on mukana useita pieniä, tilastollisesti merkitseviä tutkimuksia, meta-analyysissä voidaan helposti yliarvioida intervention teho. Julkaisuharha eli negatiivisten julkaisemattomien tutkimusten potentiaalista vaikutusta meta-analyysiin selvennetään kuvassa 4. Julkaisuharhan havaitsemiseen on olemassa sekä graafisia että tilastollisia menetelmiä, joiden tulosten tulkinta voi kuitenkin olla ongelmallista (23).

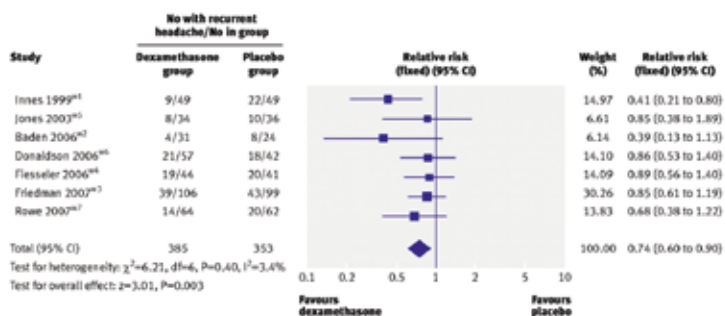
Yksittäisten tutkimusten laadun ja homogeenisuuden arviointi
Huonolaatuisten satunnaistettujen tutkimusten yhdistäminen meta-analyysin menetelmiä käyttäen ei voi johtaa tuloksiltaan luotettavaan meta-analyysiin (24). Tämän vuoksi meta-analyysiä tehtäessä onkin tärkeää arvioida analyysissä mukana olevien yksittäisten tutkimusten laatu (24, 25). Yksittäisten

satunnaistettujen tutkimusten laadun arviointiin on olemassa ainakin 25 erilaista mittaria (26). Mittareiden käyttäminen erilaisia pisteytysmenetelmiä hyödyntäen voi kuitenkin olla ongelmallista. On todettu, että yksittäisten tutkimusten laadun arviointiin käytetty mittari voi vaikeuttaa meta-analyysin tuloksiin (26). Siksi tutkimusten laatua arvioitaessa yksittäisten satunnaistettujen tutkimusten metodologiassa tulisi aina huomioida seuraavat kolme oleellista asiaa: onko satunnaistaminen tehty asianmukaisesti, onko päävastemuuttujan arviointi tehty sokkoutetusti, kuinka tutkimuksesta (syystä tai toisesta) kesken tutkimuksen pois pudonneet potilaat on huomioitu tulosten analysoinnissa (22, 25, 27). Myös muita metodologisesti tärkeitä seikkoja tutkimusten laatuun liittyen tulee huomioida tapauskohtaisesti tutkittavasta interventtiosta riippuen.

Eräs oleellinen tärkeä seikka meta-analyysin tuloksia arvioitaessa on huomioida, ovatko kaikki mukaan otettavat yksittäiset tutkimukset sellaisia, että niiden tuloksia on mielekästä yhdistää. Esimerkiksi ravitsemusta käsittelevissä tutkimuksissa osassa voi olla mukana aliravitettuja potilaita ja osassa potilaiden ravitsemustilanne voi olla normaali. Myös annettu hoito voi olla erilainen esimerkiksi lääkehoidon annostuksen tai keston perusteella. Näissä tapauksissa suoraviivainen tutkimusten yhdistäminen meta-analyysissä ei ole mielekästä yksittäisten tutkimusten heterogeenisyyden vuoksi. Heterogeenisyyden syyt tulee selvittää ja mahdollisten alaryhmien analysoiminen erillään tulee etukäteen suunnitella ja raportoida (15, 23, 28-30). Heterogeenisyyden esiintymistä meta-analyysissä voidaan myös tilastollisesti testata käyttäen esimerkiksi ns. Higginsin I²-statistiikkaa (29). Tässä testauksessa tulos ilmoitetaan prosentteina ja se voi vaihdella välillä 0-100%. Jos tulos on esimerkiksi 0%, se tarkoittaa että heterogeenisyyttä ei ole lainkaan. Jos testin I²-arvo on > 50%, tutkimusten välillä on heterogeenisyyttä ja meta-analyysin tulokset tulisi mahdollisesti esittää myös alaryhmäkohtaisesti (29, 30).

Forest plot of the effectiveness of dexamethasone compared with placebo in preventing the recurrence of acute severe migraine headache in adults.

Kuva 5



Sedgwick P BMJ 2011;342:bmj.d45

©2011 by British Medical Journal Publishing Group

BMJ

Kuva 5. Esimerkki meta-analyysin raportoinnista. Tässä on tutkittu deksametasonin tehoa äkillisen migreenikohtauksen hoidossa (28).

Meta-analyysin tulosten esittäminen ja pohdinta

Laadukkaassa meta-analyysissä tutkimuksen kulku esitetään kaaviona samaan tapaan kuin yksittäisessä satunnaistetussa kontrolloidussa tutkimuksessa (27). Kaaviossa esitetään läpikäytyjen potentiaalisesti relevanttien tutkimusten määrä, meta-analyysistä poissuljettujen tutkimusten määrä ja syyt poissulkuun. Yksittäisten satunnaistettujen tutkimusten tulokset yleensä esitetään lyhyesti (esimerkiksi otoskoko, interventio, seuranta), samoin kuin päätulos. Tulokset esitetään graafisesti yksittäisten tutkimusten osalta ja tutkimukset yhdistäen. Päävastemuuttujasta esitetään lopputuleman piste-estimaatti ja sen 95% luottamusvälit. Kuvassa 5 on esimerkki meta-analyysin päätuloksen esittämisestä graafisesti (28). Esimerkkitutkimuksessa on selvitetty parenteraalisesti kertannoksena annetun deksametasonin vaikutusta päänsäryn vähenemiseen plasebo- tai standardihoitoon verrattuna (28). Kuvassa näkyy

yksittäisten tutkimusten tulokset (piste-estimaatti ja 95% luottamusväli (CI)) ja alimpana meta-analyysin tulos (piste-estimaatti ja 95% CI) kun kaikki tutkimukset on yhdistetty. Meta-analyysin tulos on tilastollisesti merkitsevä. Kuvasta käy myös ilmi, että tilastollisessa testauksessa (ns. I²-statistiikka) tutkimusten välillä ei todettu heterogeenisyyttä (I²=3.4%) (29). Yksittäisten tutkimusten osuus koko tutkitusta potilasmää-

Laadukkaassa meta-analyysissä tutkimuksen kulku esitetään kaaviona samaan tapaan kuin yksittäisessä satunnaistetussa kontrolloidussa tutkimuksessa.

rästä on myös esitetty ("weight"). Otoskooltaan suurin tutkimus saa siis suurimman painoarvon kokonaistulosta laskettaessa.

Meta-analyysin tuloksia arvioitaessa tulee myös miettiä, voiko tulokset esittää kaikki yksittäiset

>>

tutkimukset huomioiden ja tässä yhteydessä käytetään termiä ”herkkyysanalyysi”. Tämä analyysi voi esimerkiksi paljastaa, että kokonais-tulokseen vaikuttaa huomattavasti yksi ainoa suuri tutkimus, tai että yksittäisten tutkimusten laatu voi vaihdella ja siten esimerkiksi hyvä- ja huonolaatuiset tutkimukset tulisi analysoida erikseen. Pohdintaosassa huomioidaan päätuloksen lisäksi mahdolliset harhat tutkimuksessa ja niiden merkitys tuloksille (23, 29, 30). Tulosten ulkoinen validiteetti tulisi arvioida, eli se kuinka hyvin tulokset ovat yleistettävissä kohdeväestöön (27).

Yhteenveto

On huomioitava, että meta-analyysit ovat luonteeltaan retrospektiivisiä ja osa asiantuntijoista on sitä mieltä, että tästä syystä meta-analyysit ovat enemmän tutkimushypoteeseja luovia kuin niitä testaavia.

Kuitenkin valtaosa tutkijoista pitää meta-analyysien todistusvoimaa interventioiden tehoa arvioitaessa hyvin suurena. Esimerkiksi Suomessa valtakunnallisia Käypä hoito-ohjeita laadittaessa laadukkaiden meta-analyysien todistusvoima on eri tutkimusasetelmien välisessä

hierarkiassa korkeimmalla sijalla ja näin ollen meta-analyyseillä on suuri merkitys hoidon ohjeistuksessa. Tärkeää on kuitenkin huomioida, että sekä laadukkaita että huonolaatuisia meta-analyyseja julkaistaan samoin kuin yksittäisiä satunnaistettuja tutkimuksiakin, joten julkaisun tapauskohtainen arviointi on oleellisen tärkeää. ■

Viitteet

1. Antman EM, Lau J, Kupelnick B, ym. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *Treatments for myocardial infarction*. JAMA 1992; 268: 240-8.
2. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second Internatio-

nal Study of Infarct Survival) Collaborative Group. *Lancet* 1988; 8607: 349-60.

3. GISSI. Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986; i: 397-402.
4. Lau J, Ioannidis JPA, Schmod CH. Summing up the evidence: one answer is not always enough. *Lancet* 1998; 351: 123-7.
5. Delaney A, Bagshaw SM, Ferland A, ym. A systematic review of the quality of meta-analyses in the critical care literature. *Critical Care* 2005; 9: R575-82.
6. Feinstein AR. Meta-analysis: Statistical alchemy for the 21st century. *J Clin Epidemiol* 1995; 48: 71-9.
7. Sacks HS, Berrier J, Reitman D, ym. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; 316: 450-5.
8. Smith AF. An analysis of review articles published in four anaesthesia journals. *Can J Anaesth* 1997; 44: 405-9.
9. Jadad AR, McQuay HJ. Meta-analyses to evaluate analgesic interventions: A systematic qualitative review of their methodology. *J Clin Epidemiol* 1996; 49: 235-3.
10. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. ISIS-4: A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *The Lancet* 1995; 345: 669-82.
11. Yusuf S, Teo K, Woods K. Intravenous magnesium in acute myocardial infarction. An effective, safe, simple, and inexpensive intervention. *Circulation* 1993; 87: 2043-6.
12. LeLorier J, Grégoire G, Benhaddad A, ym. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997; 337: 536-42.
13. Borzak S, Ridker PM. Discordance between meta-analyses and large-scale randomized, controlled trials. *Ann Intern Med* 1995; 123: 873-7.
14. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med* 1997; 127: 380-7.
15. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature VI. How to use an overview. *JAMA* 1994; 272: 1367-71.
16. McAuley L, Pham P, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000; 356: 1228-31.
17. Egger M, Zellweger-Zahner T, Schneider M, ym. Language bias in randomized controlled trials published in English and German. *Lancet* 1997; 350: 326-9.
18. Moher D, Fortin P, Jadada AR, ym. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996; 347: 363-6.
19. Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; 323: 101-5.
20. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; 337: 867-2.
21. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997; 315: 640-5.
22. Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998; 279: 281-6.
23. Sedgwick P. Meta-analyses: funnel plots. *BMJ* 2011; 343: d572.
24. Moher D, Pham B, Cook DJ, ym. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352: 609-13.
25. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001; 323: 42-6.
26. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282: 1054-60.
27. Moher D, Cook DJ, Eastwood S, ym. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUAROM statement. *Lancet* 1999; 354: 1896-900.
28. Sedgwick P. Meta-analysis I. *BMJ* 2011; 342: d45.
29. Sedgwick P. Meta-analyses III *BMJ* 2011; 342: d244.
30. Sedgwick P. Meta-analysis VI. *BMJ* 2011; 342: d937.