

Inleiding in evidence-based medicine

onder redactie van
M. Offringa
W.J.J. Assendelft
R.J.P.M. Scholten

5.2 Systematische reviews van interventieonderzoek¹

W.J.J. Assendelft, R.J.P.M. Scholten en L.M. Bouter

5.2.1 INLEIDING

Zoals in paragraaf 5.1 is uiteengezet, gaat een systematische review uit van een expliciete vraagstelling, een zorgvuldige, uitputtende zoekstrategie, een objectieve selectie van onderzoeken, beoordeling van de kwaliteit van de geselecteerde onderzoeken en een transparante presentatie van de resultaten. Deze opbouw geldt ook voor een systematische review van interventieonderzoek.

Idealiter is een systematische review van interventieonderzoek gebaseerd op randomised controlled trials (RCT's), omdat RCT's de minste kans op vertekening geven. Systematische reviews van niet-gerandomiseerde onderzoeken bevinden zich dan ook op een lager niveau van bewijs. Men moet zich echter realiseren dat voor veel onderwerpen en interventies RCT's simpelweg ontbreken of onmogelijk zijn. Het zou bijvoorbeeld onethisch zijn nog een gerandomiseerd onderzoek naar postexpositieprofylaxe bij hiv uit te voeren, terwijl aan de hand van retrospectief onderzoek reeds bekend is dat de relatieve risicoreductie van direct toegediend zidovudine 79% is. In dat geval kan ook een systematische review gebaseerd op observationeel onderzoek uitgevoerd worden (zie paragraaf 5.4).

Naast een kwalitatieve samenvatting van de resultaten kan een systematische review van interventieonderzoek tevens een kwantitatieve samenvatting (statistische pooling of meta-analyse) bevatten. Dit getalsmatig combineren van gegevens uit verschillende onderzoeken vergroot de precisie van de effectschattingen en maakt analyse van subgroepen mogelijk die wellicht gevoeliger zijn voor een interventie. In deze paragraaf wordt eerst aandacht besteed aan de opbouw van een systematische review van interventieonderzoek en de beoordeling

1 Voor een goed begrip van deze paragraaf dient paragraaf 4.4 bestudeerd te zijn.

van een meta-analyse. Daarna worden handvatten aangereikt voor de beoordeling van dergelijke reviews.

5.2.2 OPBOUW VAN EEN SYSTEMATISCHE REVIEW VAN INTERVENTIEONDERZOEK

Voor de beoordeling van de methodologische kwaliteit van een systematische review is het belangrijk alle relevante onderdelen van een review goed te begrijpen.

Vraagstelling

Een systematische review moet uitgaan van een expliciete klinische vraagstelling (zie hoofdstuk 2). De noodzakelijke ingrediënten van een dergelijke vraag zijn: de patiëntencategorie, de interventie, de controle-interventie, de uitkomst en eventueel ook de gewenste follow-up duur. Hiervoor wordt het Engelse acroniem PICO gebruikt: Patients, Intervention, Control, Outcome. Een voorbeeld van een vraagstelling is: wat is bij patiënten met een acute, subacute of chronische tennis-elleboog (patients) de effectiviteit, uitgedrukt in mate van ervaren herstel (outcome), van corticosteroïdinjecties (intervention) ten opzichte van een placebo of afwachtend beleid (control)? Bij deze vraag is gekozen voor ervaren herstel als uitkomst en bijvoorbeeld niet voor pijn, hoewel deze uitkomst ook in een systematische review had kunnen worden opgenomen. Vaak wordt echter voor een review een minder scherp geformuleerde vraagstelling genomen. Dan worden een of meer onderdelen van de PICO niet of minder uitgebreid uitgewerkt. De lezer moet zich in die situatie steeds afvragen welke van de conclusies van de review zijn gebaseerd op de onderdelen van de P, de I, de C en de O die de lezer bij het klinische probleem voor ogen heeft. Vraag u dus steeds af of de review niet gaat over een bredere of andere vraagstelling dan de PICO van u.

Zoekactie

Uitgaande van de vraagstelling dient op gestructureerde wijze uitgebreid (sensitief) in de internationale literatuur te zijn gezocht. Er mogen immers geen studies worden gemist. Een minimumvereiste hiervoor is een zoekactie in de elektronische databases MEDLINE, EMBASE (zie hoofdstuk 3), het Cochrane Central Register of Controlled Trials (hoofdstuk 9) en controle van de referenties van de verzamelde artikelen. Optionele aanvullende zoekacties zijn: zoeken in (gespecialiseerde) trialregisters, gedrukte indexen, het aanschrijven van experts en fabrikanten en het handmatig doorzoeken van tijdschriften.

De zoekactie dient gezien de actualiteit van een onderwerp voldoende recent te zijn. Algemene richtlijnen hiervoor zijn niet te geven, maar reviews van ouder dan vijf jaar dienen met grote voorzichtigheid gebruikt te worden. De lezer kan met de zoektermen van de review zelf in PubMed zoeken naar recentere onderzoeken en kijken of deze in lijn zijn met de conclusies van de review.

Selectie

De selectie van de in de systematische review te includeren artikelen moet bij voorkeur door ten minste twee reviewers aan de hand van expliciete in- en exclusiecriteria onafhankelijk van elkaar zijn uitgevoerd.

In veel reviews wordt de inclusie beperkt tot publicaties in de talen die de reviewers zelf machtig zijn (meestal het Engels). Het is niet duidelijk of dit in de regel tot afwijkende conclusies (bias) leidt. De impact van een taalrestrictie kan per review verschillend zijn (over- of onderschatting van het effect van de interventie).

Kwaliteitsbeoordeling

De validiteit van de aanbevelingen in een systematische review hangt sterk af van de kwaliteit van de oorspronkelijke RCT's.

Er is een veelheid aan kwaliteitsbeoordelingslijsten beschikbaar. Slechts van een klein aantal interne validiteitsitems staat vast dat deze tot vertekening kunnen leiden: geblindeerde toewijzing van de interventie ('concealment of allocation'), blinding van de effectbeoordelaars en compleetheid van de follow-up (geen selectieve uitval) (zie ook paragraaf 4.4). Een in de systematische review gebruikte kwaliteitsbeoordelingslijst dient dus ten minste deze items te bevatten. Net als de selectie dient de kwaliteitsbeoordeling in de systematische review door ten minste twee reviewers onafhankelijk van elkaar te zijn uitgevoerd.

Presentatie van de resultaten

In een systematische review dienen de gegevens die direct bij de vraagstelling aansluiten, op overzichtelijke wijze te worden gepresenteerd. De resultaten uit de oorspronkelijke RCT's worden in een voor alle RCT's identieke associatiemaat uitgedrukt. In een systematische review wordt dus geen genoegen genomen met de mededeling dat een gevonden verschil in een RCT al dan niet statistisch significant is. Voor dichotome uitkomstmaten (bijv. wel/niet genezen) wordt de associatie uitgedrukt in de oddsratio (OR), het relatieve risico (RR) of het risicoverschil (RV) met het bijbehorende 95%-betrouwbaarheidsinterval

(95%-BI), voor continue variabelen (bijv. opnameduur in dagen) in een al dan niet gestandaardiseerde verschilscore met 95%-BI. Ook de data-extractie moet bij voorkeur door ten minste twee reviewers onafhankelijk van elkaar geschieden.

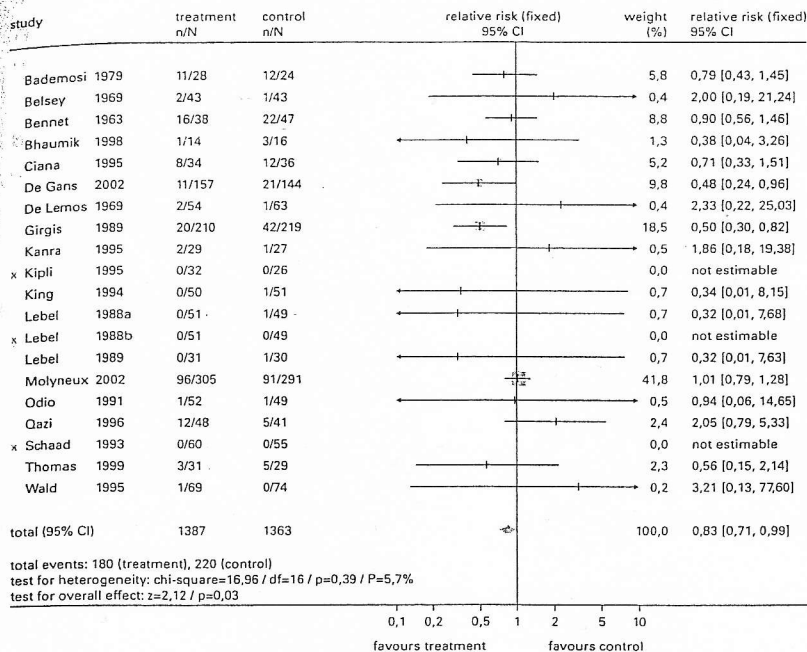
Combineren van resultaten

In de meta-analyse of statistische pooling worden de gegevens die in het kader van een systematische review zijn vergaard statistisch samengevoegd om te komen tot een schatting van het in deze onderzoeken bestudeerde effect. Daarnaast wordt natuurlijk ook gekeken naar eventuele verschillen tussen de onderzoeken onderling (heterogeniteit). Voor het combineren van effectschattingen ('pooling') zijn diverse statistische technieken beschikbaar. Met behulp van statistische pooling wordt een schatting van het effect van een interventie verkregen, die gebaseerd is op het grootst mogelijke aantal patiënten, waardoor een maximale precisie wordt verkregen (smalle betrouwbaarheidsintervallen), zodat ook kleine effecten gedetecteerd kunnen worden. Figuur 5.1 is een voorbeeld van een meta-analyse naar de effectiviteit van corticosteroiden versus placebo bij bacteriële meningitis, met overlijden als uitkomst.

Soms zijn de gebruikte meetinstrumenten voor een uitkomstmaat verschillend, zodat deze niet rechtstreeks te combineren zijn. Depressie wordt bijvoorbeeld met een groot aantal verschillende meet-schalen geëvalueerd. De uitkomsten kunnen in dat geval worden omgeschreven (gestandaardiseerd) naar zogenoemde standardised mean differences (ook wel effectsizes genoemd). Deze zijn dan wel weer te poolen. De klinische interpretatie van de uitkomsten van een dergelijke pooling is echter lastig, omdat een effectsize zich moeilijk laat terugvertalen naar een klinisch gangbare uitkomst, omdat dat er meerdere zijn. Om een gevoel te krijgen voor het gevonden effect kan de effectsize worden gekoppeld aan een standaarddeviatie in een studie. Lucassen et al. gebruikten bijvoorbeeld de standaarddeviatie van een trial in hun review (Hill et al.) om te berekenen dat de door hen in de review gevonden effectsize van 0.18 overeenkwam met een number needed to treat van 6.

Pooling is het meest in het oog springende, maar tegelijkertijd ook een veelvuldig verkeerd begrepen en misbruikt onderdeel van een systematische review. Net als bij de analyse van gegevens binnen een RCT doen statistische softwarepakketten braaf hun werk indien de gegevens per onderzoek worden ingevoerd. Voorwaarde voor pooling is echter dat de onderzoeken zowel klinisch als statistisch voldoende

review: corticosteroids for acute bacterial meningitis
 comparison: 01 all patients
 outcome: 01 mortality

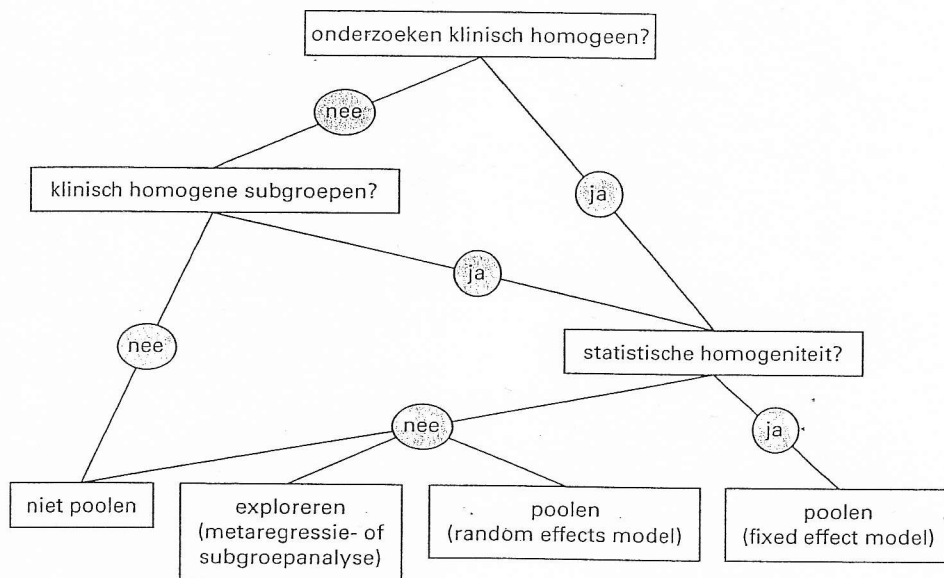


Figuur 5.1 Voorbeeld uit de Cochrane Database of Systematic Reviews van een meta-analyse. Corticosteroiden versus placebo bij acute bacteriële meningitis, met als uitkomst mortaliteit.

Ieder horizontaal lijntje geeft de uitkomst weer van een onderzoek, waarbij het lijntje het 95%-betrouwbaarheidsinterval (95%-BI) aangeeft. Het vierkantje in het midden van de lijn geeft de puntschatter van het relatieve risico (RR). Aan elk onderzoek is een gewicht gegeven. Hoe groter het onderzoek, des te groter het gewicht. Het relatieve gewicht van het onderzoek staat onder 'weight' en is ook af te lezen aan de grootte van het bij het onderzoek behorende blokje en aan de breedte van het betrouwbaarheidsinterval. De ruit is het gecombineerde (statistisch gepoolde) resultaat van de zeventien onderzoeken. De verticale lijn geeft een RR van 1 aan. Resultaten die links van de verticale lijn liggen wijzen hier op een gunstig effect van de onderzochte therapie ten opzichte van de controlegroep. De p-waarde van de test voor heterogeniteit is 0,39 en de I² 5,7%. Er lijken dus geen aanwijzingen te bestaan voor heterogeniteit. Corticosteroiden blijken volgens deze meta-analyse bij bacteriële meningitis effectiever dan een placebo: het RR is 0,83 en 95%-BI (0,71-0,99) sluit de neutrale waarde 1 uit. De kans op een slechte uitkomst neemt dus af met 17% (relatieve risicoreductie).

Bron: Van de Beek et al. Cochrane Database of Systematic Reviews 2007, Issue 1. Art. No.: CD004405.

gelijk (homogeen) zijn om bij elkaar gevoegd te mogen worden. Is dat niet het geval, dan spreekt men van heterogeniteit. Er is sprake van klinische heterogeniteit als de interventies, patiënten of de uitkomstmaten niet voldoende vergelijkbaar zijn (figuur 5.2). De beoordeling van klinische homogeniteit is bij uitstek een zaak voor de clinicus. Hiervoor zijn geen goede beoordelingscriteria beschikbaar. Klinische heterogeniteit kan worden voorkomen door vooraf (a priori) voldoende vergelijkbare subgroepen te definiëren. Indien deze niet voorhanden zijn, moet worden afgezien van pooling.



Figuur 5.2 Analyse van heterogeniteit in de resultaten van individuele onderzoeken in een systematische review.

Indien de resultaten van de verschillende onderzoeken ondanks de aanname van klinische homogeniteit (sterk) uiteenlopen, spreekt men van statistische heterogeniteit. Statistische heterogeniteit kan berusten op toeval, daadwerkelijke verschillen tussen de onderzoeken (klinische heterogeniteit), of verschillen in methodologische kwaliteit tussen onderzoeken. Vaak echter zal de oorzaak ervan onbekend zijn.

De inschatting van statistische homogeniteit is lastig. Hiervoor bestaat weliswaar een chikwadraattoets, maar aan de hand van deze relatief onvoelige statistische toets kan gemakkelijk ten onrechte gecon-

cludeerd worden dat er geen verschil is tussen de onderzoeken (homogeniteit kan niet verworpen worden). Als de meta-analyse namelijk weinig en vooral kleine onderzoeken bevat (en dat is vaak het geval), moeten de verschillen tussen de effectschattingen van de afzonderlijke onderzoeken erg groot zijn om de homogeniteittoets significant te krijgen (p -waarde $< 0,05$). Heterogeniteit wordt in dat geval gewoonweg niet opgepikt door de toets. Statistici spreken dan van een gebrek aan onderscheidingsvermogen. Door het significantieniveau van de toets iets ruimer te nemen (p -waarde $< 0,10$) kan men enigszins aan dit gebrek aan onderscheidingsvermogen tegemoetkomen. Er dient te worden opgelet bij de interpretatie van de chikwadraattoets. Een waarde groter dan de afkapwaarde (bijv. $p > 0,1$) betekent homogeniteit en beneden de afkapwaarde heterogeniteit.

Een betere, maar lastig over te brengen methode voor het bepalen van homogeniteit is visuele inspectie van de mate van overlap van de betrouwbaarheidsintervallen van de effectschattingen in de afzonderlijke onderzoeken. Als er voldoende overlap is dan is er geen sprake van statistische heterogeniteit. Formele criteria hiervoor zijn helaas niet te geven.

Naast de chikwadraattoets kan statistische heterogeniteit aan de hand van de zogenoemde I^2 statistiek gekwantificeerd worden. In een meta-analyse bestaan twee bronnen van variatie: variatie tussen personen (steekproefvariatie) en variatie tussen onderzoeken (heterogeniteit). I^2 geeft aan hoeveel procent van de totale variatie in de meta-analyse op het conto van heterogeniteit geschreven kan worden. Bij een $I^2 < 30\%$ is nauwelijks sprake van heterogeniteit en bij een $I^2 > 60\%$ van aanzienlijke heterogeniteit (in welk geval eigenlijk niet gepoold mag worden).

Bij statistische heterogeniteit kan men de heterogeniteit modelleren, verklaren of geheel afzien van pooling (zie figuur 5.2). De heterogeniteit wordt soms echter genegeerd door de statistische methode niet aan te passen (toepassing van het 'fixed-effectmodel'). Probleem hierbij is dat het fixed-effectmodel sneller significante uitkomsten geeft (te lage p -waarde of te smal betrouwbaarheidsinterval) dan het 'random-effectsmodel' (zie hierna) en dat de uitkomst dan klinisch moeilijk te interpreteren is. Door toepassing van het zogenoemde random-effectsmodel voor pooling kan met de verschillen tussen onderzoeken rekening worden gehouden. Het random-effectsmodel modelleert de aanwezige heterogeniteit, het houdt rekening met variatie tussen onderzoeken en geeft daardoor in de regel bredere betrouwbaarheidsintervallen.

Dit model is dus 'conservatiever' (geeft minder snel significante resultaten) dan het fixed-effectmodel. Een andere optie is heterogeniteit trachten te verklaren door subgroepanalyses uit te voeren. Dergelijke subgroepanalyses dienen bij voorkeur te worden gebaseerd op een aan de systematische review voorafgaand analyseplan om 'visexpedities' naar mogelijke redenen en het daarmee samenhangende risico op fout-positieve bevindingen te voorkomen. Subgroepanalyses achteraf, die niet vooraf in het protocol waren opgenomen, kunnen slechts als hypothesegenererend worden beschouwd en niet als hypothesetoetsend (bewijzend); het post-hockarakter ervan dient daarvoor in de review vermeld te worden. In veel situaties met heterogeniteit, ten slotte, is geheel afzien van pooling wellicht de beste oplossing.

Sensitiviteitsanalyses. Na de berekening van een overall effect wordt soms nagegaan hoe gevoelig (sensitief) de uitkomst is voor veranderingen in de insluitcriteria voor de meta-analyse, of wordt nagegaan of er verschil is tussen RCT's van hoge en van lage kwaliteit. Als de waarde van de gepoolde schatting hierdoor sterk verandert, dient het resultaat van de review met grote voorzichtigheid geïnterpreteerd te worden. Dit is echter geen 'verplicht' onderdeel van een systematische review.

5.2.3 BEOORDELING VAN SYSTEMATISCHE REVIEWS

Het aantal systematische reviews in de internationale literatuur stijgt exponentieel. De uitkomsten van een systematische review zijn echter pas valide en toepasbaar in de klinische praktijk indien de vraagstelling van de review voldoende aansluit bij de praktische vraagstelling van de clinicus, de methoden aansluiten bij de vraagstelling en als de methodologische kwaliteit van de systematische review voldoende is. In tabel 5.1 zijn de belangrijkste punten samengevat. Daarnaast zijn er rond de beoordeling van systematische reviews nog twee aspecten die aandacht verdienen: publicatie- en uitkomstbias en het voorkomen van tegenstrijdige systematische reviews over hetzelfde onderwerp.

Publicatie- en uitkomstbias

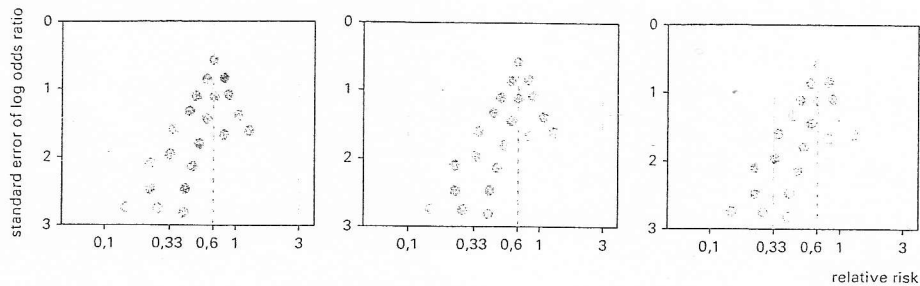
Publicatiebias is selectieve publicatie van onderzoeken. Onderzoeken waar niets uitkomt of waarin de controle-interventie het beter doet dan de experimentele interventie, worden nogal eens niet gepubliceerd. Op die manier kunnen 'positieve' onderzoeken (onderzoeken met een duidelijk gunstig effect in het voordeel van de interventie) de overhand krijgen in de review. Het resultaat is dan een overschatting van het effect van de interventie.

Tabel 5.1 Checklist voor de beoordeling van systematische reviews (checklist beschikbaar op www.pico.nu)

item	toelichting
vraagstelling	systematische review dient uit te gaan van een expliciete vraagstelling (PICO)
zoekactie	<ul style="list-style-type: none"> - ten minste MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials en controle van referenties van verzamelde artikelen - gezien actualiteit onderwerp voldoende recent
selectie	<ul style="list-style-type: none"> - bij voorkeur door twee reviewers onafhankelijk van elkaar - aan de hand van expliciete in- en exclusiecriteria - geen taalrestrictie
kwaliteitsbeoordeling	<ul style="list-style-type: none"> - bij voorkeur door twee reviewers onafhankelijk van elkaar - expliciet en daardoor reproduceerbaar voor anderen - gevalideerde checklist met ten minste de items concealment of allocation, geblindeerde uitkomstmeting en volledigheid follow-up
data-extractie en -presentatie	<ul style="list-style-type: none"> - resultaten op eenduidige wijze gepresenteerd - per studie wordt per uitkomstmaat een puntschatter van het effect en betrouwbaarheidsinterval gepresenteerd
combineren van resultaten	<ul style="list-style-type: none"> - pooling indien studies klinisch en statistisch homogeen - subgroepanalyse gebaseerd op vooraf opgesteld analyseplan
heterogeniteit	indien aanwezig beschreven en adequaat opgelost
resultaten	relevant voor de vraagstelling
conclusie	ondersteund door de resultaten

Aanwijzingen voor publicatiebias kunnen in beeld worden gebracht door middel van een 'funnelplot' (omgekeerde trechter) (zie figuur 5.3). In deze plot wordt de grootte van de effectschatting uitgezet tegen een maat voor de precisie van de effectschatting (bijvoorbeeld 1 gedeeld door de standaardfout of 1 gedeeld door de omvang van het onderzoek). Wanneer de trechter niet symmetrisch is, kan sprake zijn van publicatiebias. De vorm van een funnelplot wordt echter ook door andere factoren beïnvloed, zoals de keuze van de effectmaat, heterogeniteit en de invloed van (kleine) onderzoeken van slechte methodologische kwaliteit. Een asymmetrische funnelplot wordt dus niet altijd veroorzaakt door publicatiebias.

Een vergelijkbare en wellicht nog veel meer voorkomende vorm van publicatiebias is selectieve publicatie van uitkomsten (uitkomstbias). Uitkomsten met een niet-significant effect van de interventie worden vaak (bewust of onbewust) buiten de rapportage gelaten, wat kan leiden tot een overschot aan 'positieve' uitkomsten in de publicaties (interne publicatiebias) en in de review. Analoog aan voorgaande



Figuur 5.3 Funnel plot en publicatiebias.

Uitleg funnel plot

Links: symmetrische plot in afwezigheid van bias (de lichtgekleurde rondjes vertegenwoordigen kleine studies waarin geen positief effect werd gevonden).

Midden: asymmetrische plot door publicatiebias (kleinere studies waarin geen effect werd gevonden zijn weggelaten).

Rechts: asymmetrische plot door bias die wordt veroorzaakt door kleine studies van onvoldoende methodologische kwaliteit waarin er een overschatting is van het effect.

Donkere stippellijn is de gepoolde odds-ratio en de lichte lijn geeft de neutrale waarde 1 aan (geen effect).

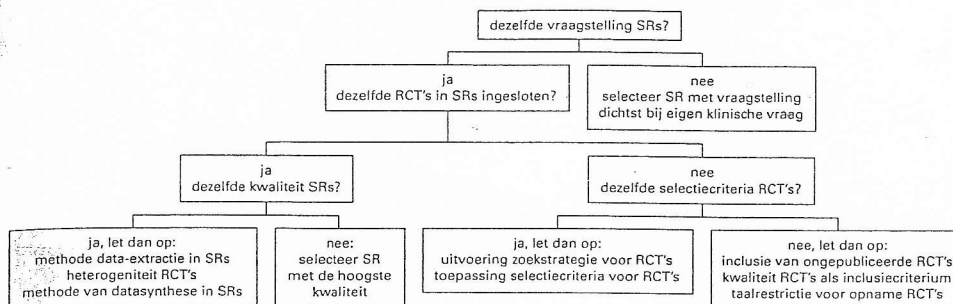
Bron: Sterne et al. BMJ 2001;323:101-5.

redenering leidt deze uitkomstbias dan tot overschatting van de daadwerkelijke effecten.

Tegenstrijdige reviews

Steeds vaker vindt de clinicus meerdere reviews over hetzelfde onderwerp. De conclusies van verschillende reviews over (schijnbaar) hetzelfde onderwerp zijn soms tegenstrijdig. De oorzaak van de verschillen kan met behulp van een algoritme worden geanalyseerd (figuur 5.4). In de regel lukt het aan de hand van dit algoritme om een of meer eenduidige systematische reviews te selecteren en tot een beslissing te komen.

Het is bij tegenstrijdige systematische reviews over hetzelfde onderwerp allereerst belangrijk om te kijken welke systematische review een vraagstelling heeft die het dichtst bij de eigen klinische vraag staat. Indien dezelfde RCT's in de verschillende systematische reviews zijn ingesloten, kan men zich richten op de methodologisch beste systematische review. Indien er geen verschil in kwaliteit is, zal de lezer een nauwkeurige analyse moeten maken van de manier waarop de gegevens zijn geëxtraheerd (bijv. door twee reviewers onafhankelijk van elkaar), hoe met heterogeniteit is omgegaan en of de uitkomsten op de



Figuur 5.4 Algoritme voor het kiezen van de best passende systematische review(s) (SR(s)) wanneer er over hetzelfde onderwerp systematische reviews met tegenstrijdige conclusies beschikbaar zijn.

Bron: Jadad et al. Can Med Assoc J 1997;156:1411-6.

juiste manier zijn gepoold. Indien niet dezelfde RCT's zijn ingesloten, dan moet allereerst worden beoordeeld of bij dezelfde vraagstelling wel dezelfde selectiecriteria voor onderzoeken zijn gehanteerd en of de reviews in actualiteit verschillen. Indien deze factoren het verschil niet verklaren, dan kan het wel of niet opnemen van ongepubliceerde onderzoeken het verschil verklaren. Daarnaast kunnen de systematische reviews onderling verschillen in een eventuele beperking op grond van taal of methodologische kwaliteit van de RCT's. Indien er met betrekking tot al deze criteria geen verschil bestaat, dan kunnen de verschillen in conclusie mogelijk worden verklaard door de verschillen in de praktische uitvoering van het zoeken en selecteren, waarbij vooral de uitvoering hiervan door twee personen onafhankelijk van elkaar menselijke fouten helpt te beperken.

5.2.4 CONCLUSIE

De systematische review verschaft op een transparante en reproduceerbare manier snel inzicht in een grote hoeveelheid informatie. Systematische reviews dienen vaak als basis voor klinische richtlijnen. Indien dergelijke richtlijnen ontbreken, is de systematische review een efficiënte en valide manier om inzicht te krijgen in een bepaald onderwerp. Het is bij een klinische vraagstelling van belang systematische reviews te selecteren die bij de vraagstelling aansluiten en van een zo hoog mogelijke methodologische kwaliteit zijn.

Literatuur

- Egger M, Davey Smith G, Altman DG. Systematic reviews in health care: meta-analysis in context. Londen: BMJ Publishing Group, 2001.
- Higgins JPT, Green S (red.). Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 [updated February 2008]. The Cochrane Collaboration, 2008 [geraadpleegd op 29 februari 2008]. Beschikbaar op www.cochrane-handbook.org.
- Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *Can Med Assoc J* 1997;156:1411-6.
- Kahn KS, Kunz R, Kleijnen J, Antes G. Systematic reviews to support evidence-based medicine. How to review and apply findings of healthcare research. Londen: Royal Society of Medicine Press Ltd, 2003.
- Montori VM, Wilczynski NL, Morgan D, Haynes RB; Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005;330:68. Epub 2004 Dec 24.
- Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, Ramsay T, Bai A, Shukla VK, Grimshaw JM. External Validation of a Measurement Tool to Assess Systematic Reviews (AMSTAR). *PLoS ONE*, 2007;2:e1350.
- Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
- Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;323:101-5.
- Tulder M van, Furlan A, Bombardier C, Bouter L; Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine* 2003;28:1290-9.