

FORMULIER Va
voor het beoordelen van een
SYSTEMATISCHE REVIEW
VAN RANDOMISED CONTROLLED TRIALS

Versie oktober 2002, geldig t/m december 2003



Evidence-Based RichtlijnOntwikkeling

Formulier Va: beoordeling systematische review van RCT's

Dit formulier is bestemd voor het beoordelen van systematische reviews van randomised controlled trials (RCT's). RCT's worden uitgevoerd ter bepaling van het effect van een therapeutische of preventieve interventie (en soms ook van een diagnostische interventie). In een systematische review wordt alle *evidence* betreffende de werkzaamheid van een interventie op een gestructureerde, overzichtelijke en reproduceerbare wijze samengevat.

Dit formulier is ontwikkeld door een werkgroep bestaande uit vertegenwoordigers van het Dutch Cochrane Centre, het Kwaliteitsinstituut voor de Gezondheidszorg CBO, het Nederlands Huisartsen Genootschap, het institute for Medical Technology Assessment, de Werkgroep Onderzoek Kwaliteit, het College voor Zorgverzekeringen, Zorgonderzoek Nederland (ZonMw) en de Orde van Medisch Specialisten en wordt ondersteund door het Nederlands Paramedisch Instituut, de Vereniging voor Integrale Kankercentra en de Werkgroep Infectieziektenpreventie.

Voor het beoordelen van de kwaliteit van andere typen onderzoek zijn eveneens formulieren ontwikkeld. Deze staan samengevat in onderstaande tabel.

Type onderzoek	Formulier
Dwarsdoorsnedeonderzoek (waarde diagnostische test)	I
Randomised controlled trial (RCT)	II
Cohortonderzoek	III
Patiënt-controleonderzoek	IV
Systematische review van RCT's (therapie en preventie)	Va
diagnostisch onderzoek	Vb
observationeel onderzoek (etiologie/"harm"/prognose)	Vc
Economische evaluatie	VI
Richtlijn	AGREE

Instructie beoordeling

- De bruikbaarheid van een publicatie voor een richtlijn wordt in de formulieren op drie facetten beoordeeld: validiteit, toepasbaarheid in de praktijk en toepasbaarheid in de Nederlandse gezondheidszorg
- Daarnaast wordt gevraagd om de belangrijkste kwantitatieve gegevens te extraheren en op een uniforme wijze te presenteren.
- De opmaak van de beoordelingsformulieren maakt het u makkelijk:
 - a) op diverse plaatsen is een beslismoment ingebouwd: indien een publicatie op dat moment niet aan de vereisten van validiteit of toepasbaarheid voldoet hoeft u met de beoordeling niet verder te gaan.
 - b) de criteria en manier van data-extractie worden telkens op de tegenoverliggende pagina kort toegelicht.

Zend opmerkingen en suggesties aangaande dit formulier naar cochrane@amc.uva.nl.

Vraag 1. *Vraagstelling.* Een systematische review dient uit te gaan van een expliciete klinische vraagstelling. De vraag dient bij voorkeur de volgende aspecten te bevatten: de aandoening, de interventie, de controlebehandeling(en), de uitkomstmaten en eventueel de gewenste follow-up duur. Een leidraad kan het Engelse acroniem "PICO" zijn: **P**atient, **I**ntervention, **C**ontrol en **O**utcome. Met name zal beoordeeld moeten worden of de conclusies van de systematische review wel gebaseerd zijn op de uitkomstmaten waarop de richtlijncommissie zich wil richten.

Vraag 2. *Zoekactie.* Uitgaande van de vraagstelling dient op gestructureerde wijze met behulp van duidelijk omschreven zoektermen in de internationale literatuur te worden gezocht. De zoektermen moeten zodanig gekozen zijn, dat naar verwachting nagenoeg alle relevante artikelen op het omschreven gebied gevonden zullen worden. Een minimum vereiste omtrent de databases is een zoekactie in MEDLINE. Indien niet in MEDLINE gezocht is, hoeft u met de beoordeling niet verder te gaan.

Een adequate zoekactie is een zoekactie in de elektronische databases van MEDLINE en EMBASE en controle van de referenties van de verzamelde artikelen. Uit onderzoek naar het zoeken naar trials is gebleken dat met een MEDLINE-zoekactie 66% van de relevante gepubliceerde artikelen gevonden wordt, terwijl met MEDLINE en EMBASE tezamen 90% werd geïdentificeerd. Ook dient het Cochrane Controlled Trials Register (CCTR) van de Cochrane Collaboration geraadpleegd te zijn. Optionele aanvullende zoekacties zijn gespecialiseerde, onderwerpspecifieke trialregisters zoals PsycINFO, CINAHL, Aidline, Toxline en Cancerlit, gedrukte indexen, het aanschrijven van experts en fabrikanten en het handmatig doorzoeken van tijdschriften.

Eveneens zullen de eventuele restricties in de zoekactie beschreven moeten zijn. Hierbij kan gedacht worden aan restrictie in jaartal van publicatie, publicatiestatus (wel of niet gepubliceerd) of in taal. Een zoektocht met taalrestricties wordt als minder valide beschouwd, omdat de uitkomst van een systematische review kan afhangen van beperkingen in taal.

Vanzelfsprekend moet de zoekactie aansluiten op de vraagstelling zoals beoordeeld in vraag 1.

Vraag 3. *Selectie.* Een uitputtende zoekactie (vraag 2) zal resulteren in een groot aantal treffers. Veel aldus gevonden artikelen zullen niet relevant zijn voor de systematische review en dus buiten de systematische review gelaten moeten worden. De selectie van artikelen die geïnccludeerd moeten worden, dient bij voorkeur door twee reviewers onafhankelijk van elkaar te geschieden aan de hand van expliciet beschreven in- en exclusiecriteria. Een minimum van twee reviewers voor de definitieve selectie is bedoeld om de kans op selectiebias in te perken.

Onder in- en exclusiecriteria vallen bijvoorbeeld: een specifieke onderzoekspopulatie (bijvoorbeeld alleen kinderen), een specifiek onderzoeksdesign (bijvoorbeeld alleen RCT's) of een specifieke interventie. De in- en exclusiecriteria dienen aan te sluiten op de vraagstelling en dienen expliciet beschreven te zijn om de reproduceerbaarheid van de systematische review te verhogen. Tevens moet de selectie vanaf de initiële database zoekactie tot de uiteindelijke inclusie in de systematische review duidelijk beschreven zijn.

Vraag 4. *Kwaliteitsbeoordeling.* De kwaliteit van de opgenomen onderzoeken dient beoordeeld te zijn. Indien geen kwaliteitsbeoordeling heeft plaatsgevonden, hoeft u niet met de beoordeling verder te gaan. Idealiter dient de kwaliteitsbeoordeling door twee reviewers onafhankelijk van elkaar aan de hand van een expliciete lijst van criteria beoordeeld te zijn. De onafhankelijke beoordeling verkleint de kans op informatiebias. De kwaliteit van de opgenomen RCT's is mede bepalend voor de validiteit van de conclusies van en aanbevelingen in de systematische review. Er zijn veel verschillende criterialijsten om de kwaliteit van RCT's te beoordelen. Slechts van een beperkt aantal criteria is door empirisch onderzoek de vertekende invloed aangetoond. Het betreft de criteria "blinding van de randomisatie" (*allocation concealment*) en blinding van de beoordelaars. De gebruikte criterialijst zal derhalve tenminste deze twee criteria moeten bevatten.

Aangezien bij *quasi-randomised trials* (ook wel *controlled clinical trials* (CCT's) genoemd) geen randomisatie plaatsvindt zal voor dit type onderzoek een criterium aan de vergelijkbaarheid van de patiëntengroepen gewijd moeten zijn.

Het resultaat van de kwaliteitsbeoordeling per individueel onderzoek zal in het artikel terug te vinden moeten zijn.

De wijze waarop tot consensus gekomen wordt en de mate van overeenstemming tussen de reviewers dient beschreven zijn.

Beoordeling van de kwaliteit van een systematische review van RCT's

Naam beoordelaar: Datum:

Titel:

Auteurs:

Bron:

Beoordeling van de validiteit

Korte beschrijving van de patiëntencategorie:

Korte beschrijving van de onderzochte interventie(s):

.....

Korte beschrijving van de controlebehandeling(en):

.....

METHODEN

1. Is de vraagstelling adequaat geformuleerd?
 Ja
 Nee
 Te weinig informatie in het artikel om dit te beantwoorden

2. Is de zoekactie adequaat uitgevoerd?
 Ja
 Nee
 Te weinig informatie in het artikel om dit te beantwoorden

3. Is de selectieprocedure van artikelen adequaat uitgevoerd?
 Ja
 Nee
 Te weinig informatie in het artikel om dit te beantwoorden

4. Is de kwaliteitsbeoordeling adequaat uitgevoerd?
 Ja
 Nee
 Te weinig informatie in het artikel om dit te beantwoorden

Vraag 5. *Data-extractie*. Data-extractie is het verzamelen van de resultaten uit de oorspronkelijke onderzoeken. Het data-extractieproces dient duidelijk beschreven te zijn en heeft weer bij voorkeur door twee reviewers onafhankelijk van elkaar plaatsgevonden.

Vraag 6. *Beschrijving oorspronkelijke onderzoeken*. De belangrijkste kenmerken van de oorspronkelijke onderzoeken moeten in de systematische review beschreven zijn. Welke items belangrijk zijn, hangt af van het onderwerp. Tenminste zullen beschreven moeten zijn:

- *onderzoeksontwerp*
- *populatie (relevante patiëntenkarakteristieken)*
- *interventie en controlebehandelingen*
- *primaire uitkomsten*
- *follow-up duur*

Vraag 7. *Adequate omgang met heterogeniteit*. Indien de oorspronkelijke onderzoeken onderling sterk verschillen is sprake van heterogeniteit. Men spreekt van klinische heterogeniteit als de patiënten, de interventies of de uitkomstmaten niet voldoende vergelijkbaar zijn. In dergelijke gevallen is het weinig zinvol de resultaten van de afzonderlijke onderzoeken te combineren tot één overall resultaat ("poolen"). Bij klinische heterogeniteit kan worden gezocht naar voldoende vergelijkbare subgroepen; indien deze niet voorhanden zijn dient te worden afgezien van pooling (zie schema op laatste pagina van dit formulier).

Hoewel in sommige situaties geen sprake is van klinische heterogeniteit, lopen de resultaten van de oorspronkelijke onderzoeken soms toch sterk uiteen. In dat geval spreekt men van statistische heterogeniteit (of "onverwachte" heterogeniteit). Bij het bestaan van statistische heterogeniteit zijn er de volgende opties (zie schema op laatste pagina):

- a. Heterogeniteit verdisconteren. Door toepassing van het zogenoemde "random effects model" wordt rekening gehouden met de variatie tussen onderzoeken (= heterogeniteit). Door deze extra variatie te verdisconteren worden bij pooling bredere betrouwbaarheidsintervallen verkregen en is dit model dus "conservatiever" (leidt minder snel tot significante resultaten dan het gangbare "fixed effect model"). Toepassing van het random-effectsmodel is alleen zinvol als sprake is van "mooie" heterogeniteit, dat wil zeggen de resultaten wijzen alle in dezelfde richting, lopen niet al te ver uiteen en de betrouwbaarheidsintervallen van de afzonderlijke onderzoeken overlappen elkaar voldoende.
- b. Heterogeniteit exploreren. Aan de hand van subgroepanalyses of metaregressie-analyse kan onderzocht worden wat de invloed is van verschillende onderzoekskenmerken (of kwaliteitskenmerken) op het resultaat. Deze *post-hoc* analyses dienen echter met grote voorzichtigheid beoordeeld te worden, aangezien er een grote kans is op het vinden van fout-positieve resultaten.
- c. Afzien van pooling. In dit geval kan bijvoorbeeld gekeken worden naar de consistentie van de resultaten (wijzen alle uitkomsten in dezelfde richting).

Vraag 8. *Statistische pooling*. Statistische pooling is het samenvoegen van de resultaten van de oorspronkelijke RCT's om te komen tot één overall schatting van het effect van de bestudeerde interventie. Voor het combineren van resultaten zijn verschillende statistische technieken beschikbaar. De voorwaarde voor pooling is dat de onderzoeken klinisch en statistisch voldoende gelijk (homogeen) zijn om bij elkaar gevoegd te worden (zie vraag 7). Indien subgroepanalyses zijn uitgevoerd, dienen deze bij voorkeur aan de hand van een vooraf opgesteld analyseplan geschied te zijn. Indien dit niet zo is dient dit te worden vermeld. Post-hoc subgroepanalyses hebben minder zeggingskracht.

Vraag 9. *Algemeen oordeel*. Hier wordt een inschatting gevraagd van de validiteit (Is dit een goede systematische review?) en toepasbaarheid (Gaat deze review over mijn PICO?). Let hierbij ook op eventuele fouten in de review die funest zijn voor de validiteit ervan (*red flags, fatal flaws*). Er zijn geen regels te geven voor welke items positief gescoord moeten worden of hoeveel items tenminste positief gescoord moeten worden. Dit is deels afhankelijk van de "state-of-the-art" met betrekking tot het betreffende onderwerp. Het gaat er hier om het samenvattend oordeel van wat de beoordelaar de werkgroep zou willen mededelen over de bruikbaarheid van het artikel voor de besluitvorming.

5. Is adequaat beschreven hoe data-extractie heeft plaatsgevonden?
 Ja
 Nee
 Te weinig informatie in het artikel om dit te beantwoorden
6. Zijn de belangrijkste kenmerken van de oorspronkelijke onderzoeken beschreven?
 Ja
 Nee
 Te weinig informatie in het artikel om dit te beantwoorden
7. Is adequaat omgegaan met klinische en statistische heterogeniteit van de onderzoeken?
 Ja
 Nee
 Te weinig informatie in het artikel om dit te beantwoorden
8. Is statistische pooling op een correcte manier uitgevoerd?
 Ja / niet van toepassing (vanwege heterogeniteit terecht geen pooling uitgevoerd)
 Nee
 Te weinig informatie in het artikel om dit te beantwoorden

ALGEMEEN OORDEEL

9. Zijn de resultaten van de systematische review valide en toepasbaar?
 Voldoende valide en toepasbaar → ga verder bij 10
 Twijfelachtig → ga verder bij 10; voorzichtig oordeel bij 13
 Onvoldoende valide en toepasbaar U kunt stoppen met het invullen van de checklist, tenzij er geen betere artikelen op dit gebied zijn (terugkoppelen naar de werkgroep)

Vraag 10. *Resultaten*. Auteurs van trials zijn soms geneigd de meest in het oog springende (significante) resultaten als belangrijkste te presenteren. Het is als beoordelaar belangrijk om zich vooraf een indruk te vormen van de klinisch of beleidsmatig meest relevante uitkomst(-en) en follow-upmoment(en). Dit zijn de belangrijkste parameters die meegenomen dienen te worden in de rapportage naar de groep. Laat u niet louter leiden door de hiërarchie van de auteurs van het artikel.

Neem, indien pooling is uitgevoerd in de systematische review, de gepresenteerde *gepoolde* resultaten over van de door u gekozen uitkomsten. (NB: U bent hier afhankelijk van de reviewers en kunt niet zelf uw eigen favoriete effectmaat berekenen, tenzij alle gegevens gerapporteerd zijn en u de hele analyse zelf kunt herhalen.)

Voor dichotome uitkomsten (genezen – niet-genezen; in leven – overleden) neemt u het relatieve risico (RR), de oddsratio (OR) of de absolute risico reductie (ARR) over.

Voor continue variabelen (numerieke waarden) neemt u de verschillscore (Engels: mean difference (MD)) over of de gestandaardiseerde verschillscore (Engels: standardised mean difference (SMD)).

Bij deze associatiematen (effectmaten) wordt een 95%-betrouwbaarheidsinterval (95%-BI) gerapporteerd. Een (gepooled) resultaat is significant indien het 95%-BI de "neutrale" waarde van de desbetreffende effectmaat niet omvat. De neutrale waarde is de waarde waarvoor geen verschil in effect bestaat tussen de onderzochte interventies. In geval van OR of RR is de neutrale waarde "1" (het geschatte effect is dus "significant" als het 95%-BI de "1" niet omvat). In geval van ARR of een (gestandaardiseerde) verschillscore is de neutrale waarde "0" (het geschatte effect is dan "significant" als het 95%-BI de "0" niet omvat).

Als het effect van verschillende interventies in dezelfde systematische review onderzocht wordt, laat u dan niet misleiden door zogenoemde indirecte vergelijkingen. Stel, dat in enkele onderzoeken in de review interventie A wordt afgezet tegen placebo en in enkele andere onderzoeken interventie B tegen placebo. De neiging bestaat dan het gepoolde effect van interventie A te vergelijken met dat van interventie B en vervolgens een uitspraak te doen over welke interventie het beste effect genereert. Dit is niet valide: doordat A en B niet direct tegen elkaar zijn afgezet, kan de invloed van bekende en onbekende prognostische factoren (zoals het risicoprofiel van de patiëntenpopulatie, setting) op de effectschatting niet onderzocht worden. Of A beter is dan B, kan alleen valide vastgesteld worden als A en B rechtstreeks in één of meerdere RCT's met elkaar zijn vergeleken.

10. Resultaten (NB: gebruik voor iedere vergelijking (contrast) een aparte pagina):

Interventie:

Controlebehandeling:

DICHOTOME UITKOMSTEN (genezen / niet-genezen; in leven / overleden)

Uitkomst:

Follow-up: weken / maanden / jaar

Aantal trials:

Associatiemaat	Waarde	95%-BI	Homogeen?
Relatieve risico (RR)			
Oddsratio (OR)			
Absolute risico reductie (ARR)			

Uitkomst:

Follow-up: weken / maanden / jaar

Aantal trials:

Associatiemaat	Waarde	95%-BI	Homogeen?
Relatieve risico (RR)			
Oddsratio (OR)			
Absolute risico reductie (ARR)			

CONTINUE UITKOMSTEN (bijvoorbeeld bloeddruk, pijnscore, kwaliteit-van-leven score)

Uitkomst:

Follow-up: weken / maanden / jaar

Aantal trials:

Associatiemaat	Waarde	95%-BI	Homogeen?
Verskil van gemiddelden (MD)			
Gestandaardiseerd verschil van gemiddelden (SMD)			

Uitkomst:

Follow-up: weken / maanden / jaar

Aantal trials:

Associatiemaat	Waarde	95%-BI	Homogeen?
Verskil van gemiddelden (MD)			
Gestandaardiseerd verschil van gemiddelden (SMD)			

Vraag 11 en 12. *Toepasbaarheid in de Nederlandse gezondheidszorg*. Beide vragen zijn een belangrijk onderdeel van richtlijnontwikkeling en dienen daarom in de werkgroep bediscussieerd te worden.

Vraag 13. *Conclusie met betrekking tot de systematische review en de waarde van de interventie(s)*

Geef hier een globale samenvatting van het eindoordeel over de systematische review. Probeer, indien aanwijzingen bestaan voor vertekening van de resultaten, tenminste een inschatting te maken van de richting van de vertekening (overschatting of onderschatting van het effect) en zo mogelijk ook over de omvang ervan. Beoordeel tevens of alle relevante uitkomsten in de review betrokken zijn (let hierbij met name op neveneffecten). Eventuele aanwijzingen voor mogelijke belangenverstremming van de auteurs met belanghebbende opdrachtgevers, kunt u hier ook rapporteren. Ook is het verstandig ingezonden brieven en/of redactionele commentaren op het hier door u beoordeelde onderzoek te raadplegen bij het formuleren van uw conclusie.

Voorbeeld: "Eindoordeel voldoende. Goed opgezet review. Er is een uitgebreide zoektocht zonder restricties uitgevoerd aan de hand van adequaat gekozen zoektermen. Er is goed omgegaan met statistische heterogeniteit. De opgenomen onderzoeken zijn voor het merendeel van lage kwaliteit. De sensitiviteitsanalyse op methodologische kwaliteit liet echter zien dat de puntschatter niet noemenswaardig beïnvloed werd. Oefentherapie lijkt effectief te zijn bij subacute en chronische rugpijn. Oefentherapie is niet effectief bij acute lage rugpijn."

TOEPASBAARHEID IN DE NEDERLANDSE GEZONDHEIDSZORG

11. Kan het gevonden resultaat worden toegepast op de Nederlandse situatie?
(hierbij valt bijvoorbeeld te denken aan de beschikbare therapeutische faciliteiten)

- Ja
- Nee
- Te weinig informatie in het artikel om dit te beantwoorden

12. Op welk(e) echelon(s) kan het resultaat worden toegepast?
(meerdere opties tegelijk mogelijk)

- algemene bevolking
- eerste lijn
- tweede lijn
 - academische ziekenhuizen
 - perifere ziekenhuizen
- derde lijn

CONCLUSIE

13. Conclusie met betrekking tot de systematische review en de waarde van de interventie(s)

SCHEMA VOOR HET OMGAAN MET HETEROGENITEIT TUSSEN ONDERZOEKEN

