FORMULIER II

voor het beoordelen van een

RANDOMISED CONTROLLED TRIAL
(RCT)

Evidence-Based RichtlijnOntwikkeling

Formulier II: beoordeling randomised controlled trial (RCT)

Dit formulier is bestemd voor het beoordelen van randomised controlled trials (RCT's). RCT's worden uitgevoerd ter bepaling van het effect van een therapeutische of preventieve interventie. Soms wordt het effect van een diagnostische interventie ook door middel van een RCT onderzocht.

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 of suggesties aangaande dit formulier naar cochrane@amc.uva.nl.

Vraag 1. *Randomisatie*. Randomisatie is een methode waarbij gebruikgemaakt wordt van het toeval om de te onderzoeken interventie en de controlebehandeling(en) toe te wijzen aan de patiënt. Randomisatie houdt in dat ieder individu (of andere eenheid van randomisatie) een gelijke kans heeft om elk van de interventies te krijgen. Een goede randomisatie kan bijvoorbeeld gebruikmaken van een tabel met aselecte (random) getallen of van een door een computer aangemaakte randomisatielijst. Er dient gewaarschuwd te worden voor andere methoden van allocatie die soms wel als randomisatie beschreven zijn, maar dit niet echt zijn: allocatie op geboortedatum, volgorde van binnenkomst, dag van de week, maand van het jaar, dossiernummer. Deze methoden heten wel “quasi random”. In dat geval is het belangrijk om extra aandacht te geven aan de vergelijkbaarheid van de groepen (vraag 6).

Vraag 2. *Blindering van de randomisatie*. Procedure waarbij wordt voorkomen dat degene die de patiënt beoordeelt en insluit op de hoogte kan zijn van de randomisatievolgorde. Goede manieren zijn: gebruik van centrale randomisatieschema’s; randomisatieschema’s die door een trial-apotheek worden beheerd; genummerde en gecodeerde verpakkingen met identieke placebo- en verum-medicatie (= werkzame medicatie); genummerde, niet-doorzichtige enveloppen; een op locatie aanwezige computer waarvan de randomisatievolgorde pas wordt vrijgegeven na opgave van de patiëntenkarakteristieken.
De in de toelichting bij vraag 1 genoemde “quasi random” procedures zijn per definitie niet blind voor randomisatie omdat degene die de patiënt in het onderzoek insluit, kan voorzien welke behandeling de patiënt zal krijgen.
Blindering van randomisatie *(concealment of allocation)* dient te worden onderscheiden van blindering van patiënten, behandelaars en effectbeoordelaars.

Vraag 3. *Blindering patiënten*. Door blindering van de patiënt wordt voorkomen dat: a) deze bewust of onbewust een grotere compliance met het protocol zal hebben, b) de uitkomstmeting door voorkeuren voor behandeling wordt beïnvloed. Blindering van de patiënt wordt bereikt door de verumbehandeling (= werkzame behandeling) en placebobehandeling identiek te maken. Medicatie moet dezelfde kleur, grootte, smaak en consistentie hebben. Ook niet-medicamenteuze placebo-interventies, zoals bijvoorbeeld fysiotherapie of ruggordels, dienen voldoende identiek te zijn om geloofwaardig over te komen. Evaluatie van het succes van blindering is gewenst, maar is voor dit item niet noodzakelijk.
Indien een onderzoek als dubbelblind wordt beschreven dient u goed na te gaan om wie het gaat: patiënt, behandelaar en/of effectbeoordelaar. Dit is op voorhand niet altijd duidelijk.

Vraag 4. *Blindering behandelaars*. Door blindering van de behandelaar wordt voorkomen dat deze, omdat hij op de hoogte is van de aard van de toegewezen behandeling: a) een bepaald enthousiasme zal uitstralen (selectieve vergroting van het placebo-effect), b) verschillende mate van adherentie aan het onderzoeksprotocol zal hebben (door bijvoorbeeld aan de placebogroep aanvullende behandeling aan te bieden). Evaluatie van het succes van blindering is gewenst, maar is voor dit item niet noodzakelijk.
Indien een onderzoek als dubbelblind wordt beschreven dient u goed na te gaan om wie het gaat: patiënt, behandelaar en/of effectbeoordelaar. Dit is op voorhand niet altijd duidelijk.

Vraag 5. *Blindering effectbeoordelaars.* Door blindering van de effectbeoordelaar wordt voorkomen dat deze de effecten van interventie en controlebehandeling verschillend zal beoordelen. Evaluatie van het succes van blindering is gewenst, maar is voor dit item niet noodzakelijk.
Indien een onderzoek als dubbelblind wordt beschreven dient u goed na te gaan om wie het gaat: patiënt, behandelaar en/of effectbeoordelaar. Dit is op voorhand niet altijd duidelijk.

Beoordeling van de kwaliteit van een randomised clinical trial (RCT)

Naam beoordelaar:.....Datum:.....

Titel: .....

Auteurs:.....

Bron: .....

Beoordeling van de validiteit

Korte beschrijving van de interventie: .....

.....

Korte beschrijving van de controlebehandeling(en): .....

.....

VALIDITEIT

1. Was de toewijzing van de interventie aan de patiënten gerandomiseerd?
   [ ] Ja
   [ ] Nee
   [ ] Te weinig informatie in het artikel om dit te beantwoorden

2. Degene die patiënten in het onderzoek insluit hoort niet op de hoogte te zijn van de randomisatievolgorde. Was dat hier het geval?
   [ ] Ja
   [ ] Nee
   [ ] Te weinig informatie in het artikel om dit te beantwoorden

3. Waren de patiënten geblindeerd voor de behandeling?
   [ ] Ja
   [ ] Nee
   [ ] Te weinig informatie in het artikel om dit te beantwoorden

4. Waren de behandelaars geblindeerd voor de behandeling?
   [ ] Ja
   [ ] Nee
   [ ] Te weinig informatie in het artikel om dit te beantwoorden

5. Waren de effectbeoordelaars geblindeerd voor de behandeling?
   [ ] Ja
   [ ] Nee
   [ ] Te weinig informatie in het artikel om dit te beantwoorden

Vraag 6. *Vergelijkbaarheid groepen.* De groepen moeten aan het begin van het onderzoek op belangrijke prognostische kenmerken voldoende gelijk zijn. Theoretisch zou alleen de toegewezen behandeling tussen de groepen verschillend moeten zijn.
Bij beoordeling kan worden gelet op:

a) Belangrijke prognostische variabelen, waaronder bijvoorbeeld ziekteduur, ernst, co-medicatie, co-morbiditeit
b) Uitgangswaarden van de belangrijkste uitkomstmaten
c) Demografische gegevens (geslacht, leeftijd)

Kleine verschillen kunnen op basis van toeval optreden. Bij grote verschillen dient beredeneerd te worden in welke mate en in welke richting de resultaten beïnvloed kunnen worden.
Er kan door de onderzoekers ook door middel van multivariate analyses gecorrigeerd zijn voor verschillen in prognostische factoren tussen de groepen.

NB: Als sprake is van *quasi randomisation* (zie vraag 1), is het belangrijk om extra aandacht te geven aan de vergelijkbaarheid van de groepen.

Vraag 7. *Loss-to-follow-up.* Het is belangrijk om per groep de aantallen patiënten bij randomisatie en bij follow-up te vergelijken. Relatief grote uitval (loss-to-follow-up) maakt een onderzoek gevoelig voor selectieve loss-to-follow-up. Aantallen en redenen voor uitval dienen gerapporteerd te zijn. Ook als er geen uitvallers waren dient dit te zijn beschreven.
Indien de redenen van uitval uit het onderzoek of de absolute aantallen uitvallers tussen de groepen verschillend zijn en tot een vertekening van de uitkomsten kunnen leiden, heet dit selectieve loss-to-follow-up.
Het is niet mogelijk om op voorhand per indicatiegebied aan te geven welk percentage loss-to-follow-up nog acceptabel is.

Vraag 8. *Intention-to-treat analyse.* Bij de analyse dient de allocatie door randomisatie gerespecteerd te worden. De patiënt hoort bij de oorspronkelijk door randomisatie gevormde groep, ongeacht eventuele co-interventies, non-compliance en dergelijke (zie vraag 9).
Naast intention-to-treat analyse kan ook nog een per-protocol analyse worden gepresenteerd. Hierbij worden alleen gegevens van patiënten gebruikt die volgens het onderzoeksprotocol zijn behandeld. Bedenk, dat een per-protocolanalyse zeer misleidend kan zijn.

Vraag 9. *Vergelijkbaarheid behandeling.* De behandeling van de patiënten in de verschillende groepen dient behalve het door randomisatie beoogde contrast geen verschillen te vertonen. Bij goed geblindeerde behandelingen is de vergelijkbaarheid van behandelingen in de regel geen probleem.
Bij de beoordeling kan worden gelet op:

a) Co-interventies. Verdeling van behandelingen anders dan de door randomisatie toegewezen. Soms worden deze door de onderzoekers onder controle en dus gelijk gehouden. In andere gevallen worden de co-interventies per groep gerapporteerd. Indien er geen melding van co-interventies wordt gemaakt dient men op de hoede te zijn.
b) Contaminatie. In geval van contaminatie krijgt of zoekt de patiënt in de loop van het onderzoek precies de behandeling die eigenlijk aan de andere groep toegewezen is.
c) Compliance. Indien de compliance met de toegewezen behandeling in de ene groep veel groter is dan in de andere kan dit de interpretatie van de gegevens verstoren.

Vraag 10. *Algemeen oordeel*. Hier wordt een inschatting van de validiteit en toepasbaarheid gevraagd. Let hierbij ook op eventuele fouten in het onderzoek 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 welk aantal 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.

6. Waren de groepen aan het begin van de trial vergelijkbaar?
   [ ] Ja
   [ ] Nee, maar in de analyses is hiervoor wel gecorrigeerd
   [ ] Nee, en in de analyses is hiervoor niet gecorrigeerd
   [ ] Te weinig informatie in het artikel om dit te beantwoorden

7. Is van een voldoende proportie van alle ingesloten patiënten een volledige follow-up beschikbaar?
   [ ] Ja
   [ ] Nee ⇐ Is selectieve loss-to-follow-up voldoende uitgesloten?
   [ ] Ja
   [ ] Nee
   [ ] Te weinig informatie in het artikel om dit te beantwoorden / loss-to-follow-up niet beschreven

8. Zijn alle ingesloten patiënten geanalyseerd in de groep waarin ze waren gerandomiseerd?
   [ ] Ja
   [ ] Nee
   [ ] Te weinig informatie in het artikel om dit te beantwoorden

9. Zijn de groepen, afgezien van de interventie, gelijk behandeld?
   [ ] Ja
   [ ] Nee
   [ ] Te weinig informatie in het artikel om dit te beantwoorden

TUSSENOORDEEL

10. Zijn de resultaten van het onderzoek valide en toepasbaar?

[ ] Voldoende valide en toepasbaar	⇐ ga verder bij 11
[ ] Twijfelachtig	⇐ ga verder bij 11
[ ] 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 11. *Resultaten*

Keuze uitkomst en follow-up duur. Auteurs zijn soms geneigd de meest in het oog springende (significante) resultaten als belangrijkste te presenteren. Het is als beoordelaar belangrijk om vooraf een indruk te vormen van de klinisch of beleidsmatig meest relevante uitkomst(en) en follow-upmoment. Dit zijn de belangrijkste parameters die meegenomen dienen te worden in de rapportage naar de groep. Pas ervoor op niet slechts op de hiërarchie van de auteurs van het artikel af te gaan.

Dichotome uitkomsten. In geval van dichotome uitkomsten (uitkomsten die slechts 2 waarden kunnen aannemen, bijvoorbeeld wel of niet genezen) kunnen verschillende associatiematen berekend worden: relatieve risico, relatieve risicoreductie, absolute risicoreductie en number needed to treat. Als de oorspronkelijke getallen gepresenteerd worden (voor notatie: zie Tabel), kan men deze associatie-maten zelf berekenen. Is dit niet het geval, dan moet men volstaan met het overnemen van de door de auteurs gepresenteerde associatiemaat (inclusief het 95%-betrouwbaarheidsinterval). Dit moet u ook doen, indien de auteurs een multivariate statistische analyse hebben uitgevoerd ter correctie voor verschillen in prognostische factoren tussen de groepen.
De formules voor het zelf berekenen van een 95%-betrouwbaarheidsinterval staan in de appendix. (Zie ook de verschillende rekenmachientjes op internet, bijv. op http://minerva.minervation.com/cebm/ of http://www.cebm.utoronto.ca/practise/ca/statscal/.)

Formules voor het berekenen van verschillende associatiematen in een RCT	Uitkomst * aanwezig	afwezig	Totaal
Interventiegroep	a	b	a + b
Controlegroep	c	d	c + d
Kans op gebeurtenis (risico) in de interventiegroep	a / (a + b)		
Kans op gebeurtenis (risico) in de controlegroep	c / (c + d)		
Absolute risico reductie (ARR)	a/(a + b) – c/(c + d)		
Number needed to treat (NNT)	1/ ARR = 1 / [| a/(a + b) – c/(c + d) |]		
Relatieve risico (RR)	[a/(a + b)] / [c/(c + d)]		
Relatieve risico reductie (RRR):			
- in geval van een ongunstige uitkomst	1 – RR		
- in geval van een gunstige uitkomst	RR – 1		

* *De uitkomst (of het eindpunt) kan zowel gewenst (bijvoorbeeld genezing) als ongewenst zijn (bijvoorbeeld bijwerking van een medicijn, overleden).*

Absolute risico reductie (ARR) = risicoverschil = verschil in absolute risico op de uitkomst tussen de interventie- en controlegroep. Indien de bestudeerde uitkomst (eindpunt) een gunstige is (genezen), wordt ook wel gesproken van een *absolute benefit increase (ABI).*
Number needed to treat (NNT) = aantal patiënten dat met de interventie behandeld dient te worden om één ongewenste gebeurtenis minder of één gewenste gebeurtenis meer te bereiken dan met de controlebehandeling verkregen zou zijn.
Relatieve risico (RR) = verhouding van absolute risico op de uitkomst tussen interventie- en controlegroep. Indien de bestudeerde uitkomst (eindpunt) een gunstige is (genezen), wordt ook wel gesproken van een *benefit ratio (BR).*
Relatieve risico reductie (RRR) = relatieve risicoverschil. In geval van een ongunstige uitkomst (bijv. overleden) en een gunstig effect van de onderzochte interventie (RR < 1 en ARR < 0) is RRR de proportionele verlaging van het risico op de slechte uitkomst (dan: RRR = 1 – RR). Bij een gunstige uitkomst (bijv. genezen) en een gunstig effect van de onderzochte interventie (RR > 1 en ARR > 0) spreekt men van “relative benefit increase” (RBI). RBI is de proportionele verhoging van het “risico” (kans) op de gunstige uitkomst (dan: RBI = RR – 1).

Continue uitkomsten. Bij continue uitkomsten wordt per behandelarm het gemiddelde effect berekend. De hier van toepassing zijnde associatiemaat is het verschil van beide gemiddelden. Voor het berekenen van een 95%-betrouwbaarheidsinterval zijn ook nog – per behandelarm – de standaard-deviatie (SD) en het aantal patiënten nodig (N). NB: Let er bij de dataextractie voor op dat de standaarddeviatie [SD] iets anders is dan de standard error (of the mean) [SE(M)]! De standaard-deviatie is de standard error of the mean maal de wortel uit het aantal patiënten in de groep. In formule: $SD = SEM * \sqrt{N}$.

11. Resultaten

In de onderstaande tabellen kunt u de meest relevante resultaten weergeven. Niet alle parameters zullen echter in het artikel vermeld staan. Deze zijn echter vaak zelf uit te rekenen met de gegevens uit het artikel (zie toelichting).

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

Uitkomst: .....

Follow-up: ..... weken / maanden / jaar

Groep	Uitkomst aanwezig	Uitkomst afwezig	Totaal
Interventiegroep			
Controlegroep			

Kans op gebeurtenis in de interventiegroep	
Kans op gebeurtenis in de controlegroep	
Absolute risico reductie (ARR)	
Number needed to treat (NNT)	
Relatieve risico (RR)	
Relatieve risico reductie (RRR)	

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

Uitkomst: .....

Follow-up: ..... weken / maanden / jaar

Groep	Gemiddelde	SD	Aantal (N)
Interventiegroep			
Controlegroep			
Verschil van gemiddelden + 95%-BI			

Vraag 12 en 13. *Toepasbaarheid in de Nederlandse gezondheidszorg.* Beide vragen zijn een belangrijk onderdeel van richtlijnontwikkeling en dienen daarom in de werkgroep bediscussieerd te worden.

Vraag 14. *Conclusie met betrekking tot het artikel en de waarde van de interventie*

Geef hier een globale samenvatting van het eindoordeel over het artikel. 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 van de interventie) en zo mogelijk ook over de grootte van de vertekening. Eventuele aanwijzingen voor mogelijke belangenverstrengeling 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 artikel. Door de aard van de interventie (oefentherapie bij lage rugpijn) is blindering van de behandelaar en patiënt vrijwel onmogelijk. Door te vergelijken met een gespreksgroep wordt echter wel goed gecorrigeerd voor aandachtseffecten. Oefentherapie lijkt effectief bij subacute en chronische lage rugpijn”.

TOEPASBAARHEID IN DE NEDERLANDSE GEZONDHEIDSZORG

12. 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

13. 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

14. Conclusie met betrekking tot het artikel en de waarde van de interventie

APPENDIX:

Formules voor het zelf berekenen van een 95%-betrouwbaarheidsinterval (95%-BI)

DICHOTOME UITKOMSTEN:

NB : op diverse internetsites zijn voor deze berekeningen ook rekenmachientjes beschikbaar
bijvoorbeeld op http://minerva.minervation.com/cebm/
of http://www.cebm.utoronto.ca/practise/ca/statscal/

Absolute risicoreductie (ARR):

$SE[ARR] = \sqrt{[ab / (a+b)^3 + cd / (c+d)^3]}$

95%-BI voor ARR: $ARR \pm 1{,}96 * SE[ARR]$

Relatieve Risico (RR) (via natuurlijke log-transformatie):

$SE[LN(RR)] = \sqrt{[1/a - 1/(a+b) + 1/c - 1/(c+d)]}$

95%-BI voor RR: $e^{LN(RR) \pm 1{,}96 * SE[LN(RR)]}$

CONTINUE UITKOMSTEN:

Verschil van gemiddelden:

$SD_P = \sqrt{[((N_I - 1) * SD^2_I + (N_C - 1) * SD^2_C) / (N_I + N_C - 2)]}$

95%-BI voor verschil van gemiddelden:

$\text{Gemiddelde}_I - \text{Gemiddelde}_C \pm t_{0,975} * SD_P * \sqrt{[1/N_I + 1/N_C]}$

I = Interventiegroep; C = Controlegroep; $t_{0,975}$ = benodigde waarde van t-verdeling met (N_I+N_C–2) vrijheidsgraden (opzoeken in tabel van t-verdelingen)