

## Gerandomiseerd klinisch onderzoek (deel 1)

Dirk Ubink · Hester Vermeulen

Dit is het eerste uit een serie van zeven artikelen over het beoordelen van diverse vormen van medisch-wetenschappelijke publicaties. De reden voor deze serie is om verpleegkundigen te helpen bij het efficiënt en kritisch beoordelen van de literatuur in hun vakgebied. Het bijhouden van de vakliteratuur is voor velen lastig omdat hiervoor eigenlijk (te) weinig tijd voor is. Daarnaast vertrouwt men vaak op de eigen ervaring, routine en klinische blik.

Toch is het goed om te realiseren dat we, ondanks onze expertise, vele malen per dag keuzes maken, zonder zeker te weten wat de juiste keuze is. Bijvoorbeeld: mag ik deze patiënt nu al wel mobiliseren? Moet ik bij iedereen postoperatief de bloeddruk meten? Heeft deze zieke patiënt zonder koorts nu wel een infectie? Moet ik altijd de huid desinfecteren voordat ik een injectie geef? Kan ik een wond ook met kraanwater reinigen?

Als u zich deze onzekerheden realiseert, kunt u deze dilemma's voortaan trachten op te lossen door een beroep te doen op bewijsmateriaal uit de vakliteratuur, zoals het EBP-principe ook bepleit. Publicaties in medische of verpleegkundige tijdschriften bevatten echter vaak heel geconcentreerde beschrijvingen in medisch of wetenschappelijk jargon van studies bij bepaalde patiënten. Het doel van deze 'toolbox' is om toch, op vrij eenvoudige wijze en zonder dat kennis van statistiek nodig is, de waarde van zulke publicaties zelf te kunnen beoordelen. Zo'n kritische beoordeling valt altijd in drie belangrijke punten uiteen. Ten eerste of de informatie in een studie valide is (dat wil zeggen: antwoord kan geven op de vraagstelling), ten tweede wat, hoe groot en hoe precies de uitkomsten van de studie zijn en ten derde of deze toepasbaar zijn in uw eigen situatie.

### Randomised Clinical Trial

In dit artikel wordt de beoordeling beschreven van een gerandomiseerd onderzoek of randomised clinical trial (RCT)<sup>1, 2</sup>, een experimenteel onderzoek waarin de effectiviteit van twee of meer behandelingen ('interventies') met elkaar wordt vergeleken. Een RCT wordt beschouwd als de beste manier van onderzoek doen naar behandeling, omdat die het sterkste bewijsmateriaal oplevert over de effectiviteit. Dit komt omdat in een RCT de onderzoeker de meeste touwtjes zelf in handen heeft: de onderzoeker bepaalt zelf welke en hoeveel patiënten meedoen, welke interventies worden uitgevoerd en welke uitkomstmaten worden gekozen. Hierdoor worden zo veel mogelijk de storende factoren uitgeschakeld die het resultaat kunnen beïnvloeden. De onderzoeker weet dan zo zeker mogelijk dat het gevonden effect ook echt toe te schrijven is aan de onderzochte behandeling.

*Als voorbeeld van een RCT zullen we een publicatie gebruiken uit het New England Journal of Medicine. 3) Hierin onderzochten de auteurs of het nodig is om bij kinderen jonger dan 3 jaar met een chronische middenoorontsteking ('loopoor') direct buisjes in hun trommelvlies te plaatsen. Gekeken werd naar het effect op de ontwikkeling van het kind qua gehoor, spraak, taal, verstandelijke ontwikkeling en gedrag. Dit werd vergeleken met een groep kinderen waarbij in principe pas vanaf een half jaar later een buisje werd geplaatst en alleen indien het nog nodig was.*

### Validiteit: de checklist

De ideale RCT heeft een aantal kenmerken. Toch zal niet elke RCT al deze kenmerken bezitten, omdat dit niet onder alle omstandigheden haalbaar is, of omdat het

niet is beschreven in de publicatie. Hierdoor laat de validiteit van het onderzoek soms te wensen over. In dat geval zullen ook de resultaten van een RCT minder of niet betrouwbaar zijn. Met een eenvoudige checklist zijn deze kenmerken na te gaan en is te beoordelen of een RCT goed (valide) is uitgevoerd en dus betrouwbare resultaten oplevert. Deze checklist is gebaseerd op het beoordelingsformulier dat het Nederlandse Cochrane Centrum beschikbaar stelt op hun website.<sup>1)</sup> We zullen de punten in deze checklist één voor één toelichten aan de hand van de voorbeeldstudie.

## 1 Randomisatie

Randomiseren is een methode waarbij de patiënten willekeurig, door het toeval, worden toegewezen aan de te onderzoeken interventie (bijvoorbeeld een nieuw verbandmateriaal) of de controlebehandeling (bijvoorbeeld het tot nog toe gebruikte verbandmateriaal). Randomisatie houdt in dat iedere patiënt een gelijke kans heeft om elk van de interventies te krijgen en voorkomt de mogelijke invloed van de voorkeur van patiënt of behandelaar. Goede manieren van randomiseren zijn bijvoorbeeld een computerprogramma dat de randomisatievolgorde pas geeft na opgave van de patiëntenkarakteristieken, gesloten en ondoorzichtige enveloppen waarin een van beide behandelingen wordt aangegeven of het opgooien van een munt.

*In de voorbeeldstudie werden 429 kinderen gerandomiseerd door middel van een computerprogramma. Dit is dus adequaat uitgevoerd.*

Checklist voor de beoordeling van RCT's

	Ja	Nee	?
1 Was de toewijzing van de interventie aan de patiënten gerandomiseerd?	X		
2 Was de onderzoeker van te voren onbekend met de randomisatievolgorde?	X		
3 Waren de patiënten geblindeerd voor de behandeling?		X	
4 Waren de behandelaars geblindeerd voor de behandeling?		X	
5 Waren de beoordelaars geblindeerd voor de behandeling?	X		
6 Waren de patiëntengroepen bij de start van de behandeling vergelijkbaar?	X		
7 Zijn uiteindelijk voldoende patiënten geanalyseerd?	X		
8 Zijn de patiënten geanalyseerd in de groep waarin ze zijn gerandomiseerd?	X		
9 Zijn de patiënten in beide groepen, afgezien van de interventie, hetzelfde behandeld?	X		

Andere methoden van toewijzing ('allocatie') zijn niet echt randomisatie te noemen: zoals allocatie op geboortedatum, volgorde van binnenkomst, dag van de week, maand van het jaar of dossiernummer. Dit zorgt namelijk niet voor een gelijke, willekeurige verdeling over beide groepen (zie ook punt 6).

## 2 Blindering van de randomisatie ('allocation concealment')

Hiermee wordt voorkomen dat degene die de patiënt randomiseert, al weet welke behandeling de volgende patiënt zal gaan krijgen. Dit is bijvoorbeeld het geval als de onderzoeker een lijst heeft waarop de randomisatievolgorde is aangegeven. Er is dan wel gerandomiseerd, maar de onderzoeker weet de volgorde en zou verleid kunnen worden deze volgorde te wijzigen, bijvoorbeeld wanneer hij vindt dat de oudere patiënt die hij nu wil randomiseren meer geschikt is voor de minder belastende behandeling A dan voor de zwaardere behandeling B. Goede manieren van allocation concealment zijn het gebruik van centrale randomisatiecentra die via de telefoon bereikbaar zijn, randomisatieschema's (een tabel met willekeurige getallen of toewijzing) die door een derde persoon worden beheerd, gecodeerde verpakkingen met identieke placebo- of werkzame medicatie, genummerde en ondoorzichtige enveloppen of een computerprogramma. De andere methoden van toewijzing genoemd bij punt 1 blinderen de randomisatie niet, omdat degene die de patiënt in het onderzoek insluit, dan kan voorzien welke behandeling de patiënt zal krijgen. NB: Blindering van de randomisatie is iets anders dan blindering van patiënten, behandelaars en effectbeoordelaars!

*In de voorbeeldstudie was allocation concealment gegarandeerd, omdat de randomisatie werd uitgevoerd met behulp van een computer door iemand die niets met de patiënt, de behandeling of de beoordeling te maken had.*

## 3 Blindering van de patiënten

Dit betekent dat de patiënt niet weet welke behandeling hij heeft gekregen. Door blindering van de patiënt wordt voorkomen dat: a) deze bewust of onbewust zich beter aan het protocol zal houden ('compliance'); b) de uitkomstmeting door voorkeur voor een bepaalde behandeling wordt beïnvloed. Blindering van de patiënt wordt bereikt door beide behandelingen identiek te maken, zoals bij medicijnstudies. Het echte en het nepmedicijn ('placebo') hebben dan dezelfde kleur, grootte, smaak en consistentie.

In de studie waren de ouders onvermijdelijk op de hoogte van de behandeling. Zij wisten immers welke behandeling hun kind had ondergaan.

#### 4 Blindering van de behandelaars

Dit betekent dat de behandelaar niet weet welke behandeling de patiënt heeft gekregen. Hierdoor wordt voorkomen dat deze, omdat hij op de hoogte is van de toegewezen behandeling:

- a. een bepaald enthousiasme zal uitstralen over een bepaalde behandeling en daardoor het oordeel van de patiënt over die behandeling positief kan beïnvloeden;
- b. aan de patiënten in één van beide groepen meer aandacht zal besteden (door bijvoorbeeld aan de placebogroep aanvullende behandeling aan te bieden).

Indien een onderzoek als ‘dubbelblind’ wordt beschreven, dient u goed na te gaan welke twee zijn geblindeerd: patiënt, behandelaar en/of effectbeoordelaar. Dit is niet altijd duidelijk.

*In de studie waren de behandelende artsen natuurlijk op de hoogte van de behandeling. Het is bij chirurgische studies vaak ondoenlijk de patiënt of de behandelaar te blinderen.*

#### 5 Blindering effectbeoordelaars

Door degene die het resultaat van de behandelingen beoordeelt te blinderen (de ‘onafhankelijke’ beoordelaar weet dan niet welke behandeling de patiënt heeft ondergaan), wordt voorkomen dat deze persoon de effecten van interventie en controlebehandeling verschillend zal beoordelen. Deze vorm van blindering is een goed alternatief indien de patiënt of behandelaar niet geblindeerd kan worden, bijvoorbeeld bij chirurgische ingrepen of het gebruik van elektrische decubitusmatrassen (immers, je hoort en ziet op welk matras iemand ligt).

*In de studie werden 3 jaar later de mogelijke effecten van de behandeling onderzocht door onderzoekers die niet wisten of het kind direct buisjes had gekregen.*

#### 6 Vergelijkbaarheid van de groepen

De kenmerken van de groepen moeten aan het begin van het onderzoek voldoende gelijk zijn. In principe zou alleen de toegewezen behandeling tussen de groepen verschillend moeten zijn. Vooral als er sprake is van oneigenlijke randomisatie (zie punt 1), is het belangrijk om extra aandacht te geven aan de vergelijkbaarheid van de groepen. Bij beoordeling hiervan kan worden gelet op:

- a. belangrijke kenmerken die de prognose kunnen beïnvloeden, bijvoorbeeld ziekte duur, ernst van de ziekte, andere medicatie, bijkomende ziekten (‘comorbiditeit’);
- b. beginwaarden van de belangrijkste uitkomstmaten;
- c. demografische gegevens (geslacht, leeftijd).

Als bijvoorbeeld de ene groep is samengesteld uit patiënten met meer comorbiditeit of meer oudere patiënten, kan het zijn dat dááruit het verschil verklaard kan worden en niet door het verschil tussen de onderzochte interventies.

*In de studie bleken beide groepen kinderen vergelijkbaar voor wat betreft hun woonplaats (stad of platteland), geslacht, etniciteit en opleiding van de ouders. Dit zijn relevante factoren die van invloed zouden kunnen zijn op het beloop van de middenoorontsteking.*

#### 7 Loss to follow-up

Als in een van beide groepen veel patiënten gedurende de behandel- en vervolperiode zijn uitgevallen (‘loss to follow-up’) en er dus minder patiënten uiteindelijk zijn geanalyseerd, moet dit zijn gerapporteerd. Het kan immers zijn dat juist de ziekste patiënten zijn uitgevallen (en het gemeten behandelresultaat bij de overigen dus overdreven positief is) of juist de genezen patiënten zijn weggebleven. Als vuistregel wordt een follow-up van 80 procent of meer van de patiënten als voldoende beschouwd.

*In de studie bleken 201 van de 216 kinderen die de directe behandeling hadden geloot ook geanalyseerd te zijn op hun zesde levensjaar. Dat is dus een prima follow-up van 97 procent. Voor de uitgestelde behandeling was de followup 91 procent (194 van de 213).*

#### 8 Intention-to-treat analyse

De patiënten dienen geanalyseerd te worden in de behandelgroep waarin ze gerandomiseerd zijn (‘intention to treat’), ongeacht eventuele verandering van de behandeling, overstappen naar de andere behandelgroep (‘cross-over’) of het stoppen met de behandeling (‘noncompliance’).

*In de voorbeeldstudie wordt gemeld dat het intention-to-treat principe werd aangehouden. Dit blijkt ook uit de vermelding dat uiteindelijk 184 van de 216 (85%) kinderen een buisje hadden gekregen in de directe groep. Vijftien procent had dus wel de directe behandeling geloot, maar niet gekregen. Toch werden ze in die groep geanalyseerd, omdat dat het aanvankelijke behandelplan was. In de uitgestelde groep hadden overigens toch nog 87 van de 213 (41%) een buisje gekregen.*

Naast de intention-to-treat analyse kan ook nog een per-protocolanalyse worden gepresenteerd. Hierbij worden alleen gegevens van patiënten gebruikt die daadwerkelijk de behandeling volgens het onderzoeksprotocol hebben gehad.

*In dit voorbeeld zouden voor een per-protocolanalyse de 159 kinderen in de directe groep plus de 21 in de uitgestelde groep die toch een buisje hebben gehad, bij elkaar genomen moeten worden om te vergelijken met de overige kinderen zonder buisje.*

Bedenk dat een per-protocolanalyse misleidend kan zijn, omdat hiermee de randomisatie teniet wordt gedaan. Dit kan weer leiden tot overschatting van het effect.

## 9 Vergelijkbaarheid van de behandeling

De behandeling van de patiënten mag alleen verschillen in de door randomisatie aangegeven behandeling.

*Als men bijvoorbeeld antibiotica zou geven aan de kinderen zonder buisjes, maar niet aan de kinderen met buisjes, is een verschillend eindresultaat niet (alleen) aan de trommelvliesbuisjes toe te schrijven. Dit lijkt in deze studie niet het geval geweest te zijn.*

Door het beantwoorden van bovengenoemde punten krijgt u een indruk hoe degelijk ('valide') de studie is uitgevoerd. Als er minpunten zijn gevonden, moet u zich afvragen of de uitvoering van het onderzoek beter had gekund.

*Bijvoorbeeld: de blinding van behandelende artsen en de ouders van de kinderen in deze studie is logischerwijs onmogelijk. Dit is dan nog wel acceptabel. Maar als bijvoorbeeld een groot aantal patiënten is uitgevallen, zijn de resultaten onbetrouwbaar!*

Zijn er ernstige tekortkomingen gevonden in de validiteit die wel beter hadden gekund en gemoeten, dan zijn de resultaten vals of vertekend en dus niet betrouwbaar. In dat geval hoeft men niet verder te lezen. Het is dan beter om een andere publicatie te zoeken die misschien niet evenveel bewijskracht heeft, maar wel goed en betrouwbaar is uitgevoerd. Zijn er geen (ernstige) tekortkomingen gevonden in de validiteit, zoals in deze studie het geval is, dan kunnen de resultaten veilig worden geïnterpreteerd.

## Resultaten

Auteurs van artikelen zijn soms geneigd de meest in het oog springende (significante) resultaten als belangrijkste te presenteren. Het is daarom als lezer belangrijk om vooraf te bedenken welke uitkomst(en) voor de patiënt het meest relevant zijn. In de voorbeeldstudie hebben de

auteurs van tevoren de mogelijke effecten op gehoor, spraak, taal en ontwikkeling terecht als relevant aange merkt. Van deze parameters willen we de resultaten dus ook zeker weten. Pas er echter voor op om niet alleen af te gaan op de informatie en conclusies van de auteurs! Doordat er bijvoorbeeld een niet-relevante uitkomst wel als positief wordt gerapporteerd, beïnvloedt dat je denkrichting.

Als een uitkomstmaat slechts twee waarden kan hebben (bijvoorbeeld jongen/meisje, wel of niet verbeterd) kunnen de resultaten het mooist in een 2 x 2-tabel worden weergegeven. Hiermee is dan goed het effect van de behandelingen te vergelijken. Dit effect wordt vaak uitgedrukt als 'relatief risico', 'absolute risicoreductie' of 'number needed to treat'. De berekening en uitleg hiervan wordt weergegeven in tabel 1. In ons voorbeeld is de kans op gehoorschade in beide behandelgroepen vergeleken. Er blijkt 3,8 procent meer gehoorschade te zijn in de groep die direct een buisje heeft gekregen. Als de oorspronkelijke getallen niet zouden zijn vermeld, moet men genoeg nemen met de resultaten die de auteurs beschrijven. Om de precisie van het resultaat weer te geven wordt soms het 95%-betrouwbaarheidsinterval erbij weergegeven. Dit geeft aan dat in 95 procent van de gevallen het echte resultaat tussen deze twee waarden ligt. Hoe smaller dit betrouwbaarheidsinterval, des te betrouwbaarder de uitkomst.

*In het voorbeeld heeft het gevonden verschil van 3,8 procent een 95%-betrouwbaarheidsinterval dat loopt van -1,4% tot +9,1%. Dit betekent dat het direct plaatsen van trommelvliesbuisjes geen duidelijk gunstig of ongunstig effect heeft op het gehoor; het zou immers 1,4% minder, maar ook 9,1% meer gehoorschade kunnen veroorzaken!*

Ook kan een p-waarde bij dit verschil worden genoemd. Een waarde kleiner dan 0,05 beschouwt men als 'statistisch significant'. Dit betekent dat de kans dat het gevonden verschil slechts op toeval berust minder is dan 5 procent. Als de uitkomst een continue maat is (bijvoorbeeld leeftijd, woordenschat of IQ-score) kan geen 2 x 2-tabel worden gemaakt, maar wordt voor elke behandeling het gemiddelde effect berekend en kijkt men naar het verschil van beide gemiddelden. Ook hierbij wordt de precisie weergegeven met een 95%-betrouwbaarheidsinterval of een standaarddeviatie (SD).

Kinderen die direct een buisje hadden gekregen kenden op hun zesde levensjaar gemiddeld 183 verschillende woorden (SD: 36), kinderen in de andere groep gemiddeld 175 woorden (SD: 36). Het verschil tussen beide gemiddelden had een betrouwbaarheidsinterval van 0,0 tot 14,4 woorden. Dit toont dus een minimaal verschil ten gunste van de kinderen met direct geplaatste buisjes.

Ook hier geeft een eventuele p-waarde bij het verschil tussen beide behandelgroepen informatie over de

Tabel 1. Formules voor het berekenen van verschillende uitkomstmaten in een RCT

	aanwezig	afwezig	afwijkend	normaal
Interventiegroep (direct een buisje)	a	b	19	180
Controlegroep (evt. later een buisje)	c	d	11	181
	Uitkomst			Gehoortest *
Kans op gebeurtenis (risico) in de interventiegroep	a / (a + b)			9,5%
Kans op gebeurtenis (risico) in de controlegroep	c / (c + d)			5,7%
Absolute risicoreductie (ARR)	a/(a + b) – c/(c + d)			3,8%
Number needed to treat (NNT)	1/ ARR			26
Relatief risico (RR)	[ a/(a + b) ] / [ c/(c + d) ]			1,67

\* Niet bij alle gerandomiseerde kinderen is een gehoortest uitgevoerd

Absolute risico reductie (ARR) = risicoverschil = verschil in risico op de uitkomst tussen de interventie- en controlegroep.

Number needed to treat (NNT) = aantal patiënten dat met de interventie behandeld dient te worden om één ongewenste gebeurtenis (gehoorsverlies) minder of één gewenste gebeurtenis meer te bereiken dan met de controlebehandeling verkregen zou zijn.

Relatief risico (RR) = verhouding van het risico op de uitkomst tussen de interventie- en controlegroep.

statistische significantie (in dit geval was de p-waarde precies 0,05).

## Toepasbaarheid

De laatste stap is het beoordelen of de gevonden resultaten ook toepasbaar zijn in de eigen situatie ('generaliseerbaarheid'). Dit hangt in de eerste plaats af van de beschikbaarheid (bijvoorbeeld als het gaat om nieuwe materialen), de aanwezige ervaring (bijvoorbeeld als het een nieuwe techniek betreft) en de kosten van de nieuwe behandeling in de eigen setting.

*Als het gaat om trommelvliesbuisjes is beschikbaarheid en ervaring geen probleem; het is immers een veel toegepaste behandeling.*

Ten tweede: zijn de patiënten die in deze studie zijn onderzocht wel vergelijkbaar met die in uw eigen situatie? Denk aan leeftijd, geslacht, ernst van de ziekte. Maar ook: gaat het om een academische patiëntenpopulatie of patiënten in de wijk? Dit bepaalt natuurlijk in hoeverre de gevonden resultaten ook bij uw patiënt toepasbaar zijn.

*In deze studie waren het verder gezonde kinderen in de praktijk van huisartsen, kinder- en KNO-artsen. Wel moet gezegd worden dat de kinderen in deze studie een chronische middenoorontsteking hadden; dus zijn deze resultaten niet automatisch toepasbaar op kinderen die herhaalde, acute middenoorontstekingen hebben (gehad) of een andere handicap. In dergelijke gevallen kan bijvoorbeeld ernstige pijn toch de doorslag geven om een buisje te plaatsen. Ten slotte is het belangrijk om de voordelen af te wegen tegen de nadelen: als een operatieve ingreep wordt*

*overwogen, moet men die afwegen tegen de risico's die bij elke operatie kunnen optreden. En als een behandeling effectief lijkt, moet men nagaan of het ook wel betaalbaar is.*

Als de resultaten van de studie inderdaad toepasbaar blijken, moet men ook de bereidheid hebben om deze nieuwe behandeling in te voeren en dus het huidige beleid te veranderen. Dit is soms erg lastig. Het schrijven van een evidence-based protocol of het wijzigen van de richtlijn wil immers niet zeggen dat het ook wordt uitgevoerd. Het veranderen van oude gewoontes is voor de meeste mensen nu eenmaal niet gemakkelijk. Zeker als het een belangrijk deel zou zijn van de werkzaamheden (en inkomsten!) van KNO-artsen. Het implementeren van evidence is dan weer een heel andere tak van sport. Globaal geldt dat de implementatie het beste lukt als men op zo veel mogelijk verschillende manieren deze nieuwe evidence of richtlijn onder de aandacht weet te brengen van de gebruikers.

*De resultaten van deze studie hebben wel – naast een flinke discussie in medische tijdschriften – geleid tot een richtlijn voor huisartsen, kinder- en KNO-artsen in de Verenigde Staten, waarin het advies is om bij kinderen met een loopoor nog 3 tot 6 maanden af te wachten, tenzij er bij controle duidelijk gehoorverlies bestaat of afwijkingen aan trommelvlies of middenoor.*

## Samenvatting

Een RCT is een van de beste manieren van onderzoek, omdat deze het sterkste bewijsmateriaal oplevert over de effectiviteit van een behandeling. Toch is het verstandig

om te beoordelen of de studie goed is uitgevoerd en of de resultaten in de eigen praktijk toepasbaar zijn. Met de checklist in dit artikel kan deze validiteit van een RCT eenvoudig worden beoordeeld. Alleen al door het beoordelen van de uitvoering van het onderzoek kunt u bepalen of de resultaten betrouwbaar zijn. Op deze wijze kunt u gericht, op basis van een vraag of onzekerheid die u zelf hebt gesignaleerd in de praktijk, een antwoord vinden dat weer in uw eigen situatie toepasbaar is. Dit kan een hoop leestijd schelen en zo houdt u heel praktisch uw vakkennis bij!

## Literatuur

- 1) [www.cochrane.nl](http://www.cochrane.nl)
- 2) Offringa M, Assendelft WJJ, Scholten RJPM. Inleiding in evidence-based medicine. Bohn Stafleu Van Loghum. Houten 2003.
- 3) Paradise JL, Feldman HM, Campbell TF, Dollaghan CA, Colborn DK, Bernard BS, Rockette HE, Janosky JE, Pitcairn DL, Sabo DL, Kurs-Lasky M, Smith CG. Effect of early or delayed insertion of tympanostomy tubes for persistent otitis media on developmental outcomes at the age of three years. *NEJM* 2005;353(6):576-586.