

# Evidence-based tandheelkunde en het beoordelen van de waarde en toepasbaarheid van onderzoek

Om als tandarts-algemeen practicus een geldig antwoord te kunnen geven op een vraag gesteld volgens de 'probleem-interventie-comparison-outcome' (PICO)-systematiek, kunnen vragen worden gesteld die de validiteit, het belang en de toepasbaarheid van een onderzoek betreffen. De validiteit wordt bepaald door de plek van het onderzoek op de ladder van 'evidence', en door een methodologische kwaliteitsbeoordeling, waarbij de uitvoering van het onderzoek wordt beoordeeld op basis van een aantal criteria. Een gerandomiseerd gecontroleerd klinisch onderzoek kan beoordeeld worden op randomisatieprocedure, blinding van patiënten, effectbeoordelaars en behandelaars, en uitval. Bij cohortonderzoek en patiënt-controleonderzoek zijn de criteria waarop een artikel beoordeeld kan worden vergelijkbaarheid van de groepen, onafhankelijkheid van de metingen en duur en volledigheid van de follow-up. Bij alle soorten onderzoek is het daarnaast van belang dat de kwaliteit van de meetmethoden goed is. De beoordelingscriteria voor systematisch literatuuronderzoek zijn een heldere vraagstelling, een duidelijk beschreven en te dupliceren zoekactie van artikelen en een onafhankelijke beoordeling van de methodologische kwaliteit van de artikelen. Ten slotte moet de tandarts-algemeen practicus bij het bepalen van de toepasbaarheid letten op de steekproef en het type patiënt dat is gebruikt in het onderzoek.

Aartman IHA, Loveren C van. Evidence-based tandheelkunde en het beoordelen van de waarde en toepasbaarheid van onderzoek. Ned Tijdschr Tandheelkd 2007; 114: 166-171

## Inleiding

In het proces van 'evidence-based' tandheelkunde, dat start met het vertalen van het klinische probleem in een beantwoordbare vraag, volgt na het zoeken naar en het vinden van wetenschappelijke artikelen de fase van het beoordelen van de informatie op methodologische kwaliteit en toepasbaarheid (Offringa et al, 2000). Om een geldig antwoord te kunnen geven op een vraag gesteld volgens de 'probleem-interventie-comparison-outcome' (PICO)-systematiek, dienen er 3 vragen te worden beantwoord (Forrest en Miller, 2003):

1. Zijn de resultaten van het onderzoek valide, geldig?
2. Wat zijn de resultaten en wat is het belang ervan?
3. Helpen deze resultaten mijn patiënten; is de informatie van toepassing?

Het antwoord op de eerste vraag wordt deels bepaald door de plaats van het gebruikte onderzoeksontwerp op de zogenaamde ladder van 'evidence'. Het artikel van Aartman en Van Loveren (2007) elders in dit themanummer behandelt de onderzoeksontwerpen en de mate van interne validiteit van deze ontwerpen. Daarnaast wordt het antwoord op de eerste vraag bepaald door na te gaan in hoeverre de uitvoering van het onderzoek voldoet aan de specifieke eisen die worden gesteld aan het ontwerp. Een onderzoek met een

ontwerp dat veel zeggingskracht heeft en dat goed is uitgevoerd op basis van toetsing aan de gestelde criteria, heeft meer zeggingskracht dan een onderzoek dat minder zorgvuldig is uitgevoerd of een ontwerp heeft dat lager staat op de ladder van evidence. In dit artikel zullen beoordelingscriteria worden gepresenteerd aan de hand waarvan de uitvoering van gerandomiseerde gecontroleerde onderzoeken ('randomised controlled trials', RCT), cohortonderzoeken, patiënt-controleonderzoeken en systematische literatuuronderzoeken kritisch kunnen worden beschouwd om zo de vraag te kunnen beantwoorden of de resultaten van het onderzoek valide zijn. Ten slotte komt in deze bijdrage de vraag naar de toepasbaarheid van de informatie aan de orde. Aan de vraag naar de resultaten en het belang ervan zal in dit artikel geen aandacht worden besteed. Niet omdat het niet belangrijk genoeg is, in tegendeel, maar er is in deze bijdrage gekozen voor de nadruk op de validiteit van de geselecteerde wetenschappelijke artikelen. Voor een bespreking van het belang van de resultaten, waarbij begrippen als statistische significantie, klinische relevantie, en grootte en precisie van het effect aan de orde komen, wordt verwezen naar andere bronnen, zoals Offringa et al (2000), Kuiper et al (2004) en Hackshaw et al (2006).

## Beoordelingscriteria voor gerandomiseerd gecontroleerd onderzoek

Een gerandomiseerd gecontroleerd onderzoek heeft van de oorspronkelijke vormen van onderzoek de hoogste interne validiteit, maar de resultaten kunnen meer of minder valide zijn, afhankelijk van de daadwerkelijke uitvoering. De validiteit van een gerandomiseerd gecontroleerd onderzoek kan worden beoordeeld aan de hand van onderstaande aspecten, die ook worden genoemd in tabel 1. Voor al de te beoordelen punten geldt dat als eraan voldaan is, dit ten goede komt aan de validiteit. Wanneer niet aan de voorwaarden wordt voldaan, vermindert de interne validiteit van het onderzoek.

### At random toewijzing

Ten eerste zal bij het onderzoek moeten worden bepaald of at random toewijzing plaatsvond. Is daar geen sprake van, dan betreft het geen gerandomiseerd gecontroleerd onderzoek maar een cohortonderzoek, patiënt-controleonderzoek of anderszins. Wanneer er sprake is van at random toewijzing, moet worden beoordeeld of deze goed genoeg gedaan is. Er mag immers geen sprake zijn van enige vertekening of selectie. Een toewijzing om en om, bijvoorbeeld, kan niet als at random worden beschouwd. Een dergelijke systematiek in de toewijzing brengt het gevaar met zich mee dat in groep A een bepaald type patiënt is oververtegenwoordigd ten opzichte van in groep B, doordat bijvoorbeeld de jongere patiënten op de oneven plekken op de wachtlijst zijn gezet. Dit lijkt misschien vergezocht, maar het gaat erom dat dit soort vertekening uitgesloten moet zijn. Een ander voorbeeld van een onvoldoende at random toewijzing is wanneer patiënten die zich 's ochtends melden, aan de éne groep worden toegewezen en patiënten die zich 's middags melden, aan de andere. Het is dus ook belangrijk dat degene die de patiënten selecteert voor deelname aan het onderzoek, niet op de hoogte is van de randomisatievolgorde. Dit wordt ook wel 'concealment of allocation' genoemd.

Hoe kan de tandarts-algemeen practicus beoordelen of de randomisatie goed is uitgevoerd? Idealiter zou dit in elk artikel nauwkeurig beschreven moeten zijn. Goede randomisatie kan bijvoorbeeld gebeuren met behulp van enveloppen met at random nummers of met een computerprogramma. Is dat het geval, dan kan worden geconcludeerd dat het wel goed zit met de randomisatie. Blijkt de gebruikte methode vertekening niet uit te sluiten, dan scoort een onderzoek negatief op dit aspect. Jammer genoeg zal ook regelmatig moeten worden geconcludeerd dat de manier van toewijzing niet te beoordelen is, omdat informatie hierover ontbreekt. De beoordeling van dit punt kan dan ook worden genoteerd als niet te beoordelen/onbekend en betekent een (mogelijk) negatieve score (zie tab. 1).

Tegenwoordig wordt aan auteurs van artikelen waarin een gerandomiseerd gecontroleerd onderzoek wordt beschreven ook wel gevraagd dit te doen aan de hand van de zoge-

naamde 'CONSORT guidelines', waarin onder andere is opgenomen dat de manier van randomiseren moet worden beschreven. CONSORT staat voor 'consolidated standards of reporting trials' en deze richtlijn is ontwikkeld door een internationale groep van onderzoekers en tijdschriftredacteuren (Moher et al, 2003). De richtlijn is te vinden via de internetpagina [www.consort-statement.org](http://www.consort-statement.org).

### Gelijkheid van de groepen

Waar bij een gerandomiseerd gecontroleerd onderzoek ook op gelet moet worden, is dat er daadwerkelijk maar 1 verschil aanwezig is tussen de groepen. Is het door de onderzoeker aangebrachte gewenste verschil ook echt het enige verschil? Door de at random toewijzing dient men ervan uit te gaan dat de groepen bij aanvang gelijk zijn. In sommige onderzoeken wordt dit vooraf gecontroleerd door de groepen te vergelijken met betrekking tot een aantal relevante variabelen als leeftijd, sekse en doorvoormetingen.

### Eenduidige interventie

Het zou ook kunnen dat de onderzoeker onbedoeld meer variatie tussen de groepen aanbrengt dan gewenst, bijvoorbeeld door niet alleen het type tandenborstel te variëren, maar ook de poetstijd. Als er dan een verschil is achteraf, is niet ondubbelzinnig vast te stellen dat het aan het type tandenborstel ligt. Alles moet zoveel mogelijk constant worden gehouden, hetzelfde zijn in de 2 groepen, om de causale uitspraak mogelijk te maken.

### Uitval

In een onderzoek kan het om allerlei redenen voorkomen dat patiënten halverwege uitvallen. Deze uitval moet zoveel mogelijk tot een minimum zijn beperkt, maar vooral niet (veel) vaker voorkomen in de éne groep ten opzichte van de andere, of in de éne groep om andere redenen dan in de andere. Als dat het geval is, kan het zijn dat de uitval selectief is, waardoor ongewenste verschillen tussen de groepen ontstaan. De redenen van uitval moeten dan ook zijn beschreven in het artikel.

### Blinding van de patiënt

Indien mogelijk zou de patiënt in een onderzoek niet op de hoogte moeten zijn van de behandelconditie waarin hij verkeert. Wanneer de patiënt weet in welke conditie hij of zij zit, kan dit de score op de uitkomstmaat beïnvloeden doordat de patiënt ook ideeën heeft over wat bijvoorbeeld een betere behandelingsmethode is. Blinding van patiënten is in veel tandheelkundig onderzoek niet mogelijk, neem bijvoorbeeld onderzoek waarin handmatig met elektrisch poetsen wordt vergeleken, of scaling en root planing versus antibiotica bij parodontitispatiënten. Voorbeelden van onderzoek waar het wel goed mogelijk is, zijn onderzoek waarin tandpasta's van verschillende samenstellingen worden vergeleken, onderzoek waarin verschillende typen com-

posiet worden gebruikt en onderzoek waarin men verschillende concentraties van een mondspoelmiddel vergelijkt. De wijze van blinding moet duidelijk zijn beschreven in het artikel.

#### **Blinding van de effectbeoordelaar**

Bij de meting van de uitkomstmaat, de effectbeoordeling, zou de beoordelaar niet moeten weten in welke onderzoeksgroep de patiënt zit. Als er niet geblindeerd is, bestaat het gevaar dat de meting van de uitkomstmaat niet op dezelfde manier is gedaan, en dat er daardoor verschillen tussen de groepen zijn gevonden. Dit kan gebeuren doordat de beoordelaar een verwachting over de uitkomst van het onderzoek heeft of een persoonlijke voorkeur voor de éne behandeling ten opzichte van de andere. Wederom, daar waar mogelijk moet de beoordeling van de uitkomstmaat onafhankelijk gebeuren van de behandeling.

Ook indien de metingen niet door personen worden verricht, dient te worden gecontroleerd of de metingen in beide condities op dezelfde manier zijn uitgevoerd. Dat wil zeggen, waren de omstandigheden hetzelfde, eventueel het tijdstip van de dag en dergelijke.

#### **Blinding van de behandelaar**

Er wordt vaak gesproken over dubbelblind onderzoek, maar idealiter zou een onderzoek zelfs tripelblind moeten zijn uitgevoerd. Dat wil zeggen dat, indien mogelijk, de behandelaar ook niet op de hoogte mag zijn van de soort behandeling, de experimentele of de controle, van de proefpersoon. Van de verwachtingen of wensen van de behandelaar zou immers een ongewenst effect kunnen uitgaan. Het spreekt voor zich dat blinding van de behandelaar doorgaans onmogelijk is, zoals bij het vergelijken van 2 soorten vulmateriaal of van een gewone versus een overkappingsprothese. Maar bij het vergelijken van handmatig met elektrisch poetsen zou men ervoor kunnen zorgen dat de instructie onafhankelijk van de kennis over het type tandenborstel wordt gegeven. Bij de vergelijking van 2 soorten antibiotica kan men ervoor zorgen dat degene die het medicijn uitreikt en de toelichting geeft, ook niet kan zien om welk middel het gaat. Alleen een onafhankelijk persoon, die niet direct met de patiënt contact heeft, is dan op de hoogte van de groep waarin de patiënt zit.

#### **Kwaliteit meetmethoden**

Een laatste aspect waarover een oordeel moet worden geveld, betreft de kwaliteit van de gebruikte meetinstrumenten. De gebruikte meetmethoden moeten valide en betrouwbaar zijn. Met een valide meetinstrument wordt gemeten wat de onderzoekers poogden te meten. Met andere woorden, de meetwaarden die het oplevert zijn geldig en relevant. Er staan de onderzoeker verschillende methoden ter beschikking om de validiteit van nieuwe meetinstrumenten te bepalen, zoals het bepalen van de sensitiviteit en de specificiteit (zie Mileman en Van den Hout, 2007). Het

artikel moet dan ook door de lezer worden gescreend op informatie over dit onderwerp, waarna kan worden beoordeeld of dit voldoende is.

Een betrouwbaar meetinstrument meet consistent. De scores op de uitkomstmaat zouden niet moeten fluctueren, al naar gelang het moment van meten, de persoon die de meting verricht enzovoorts. Idealiter zou zowel de inter- als de intrabeoordelaarsbetrouwbaarheid moeten zijn gegeven, berekend bijvoorbeeld met behulp van Cohen's kappa (Hunt, 1986). Kortom, van belang voor de tandarts-algemeen practicus is dat er in het artikel informatie wordt gegeven over de validiteit en de betrouwbaarheid van de gebruikte meetmethoden om de uitkomstmaten te bepalen. Indien geconcludeerd moeten worden dat de validiteit en/of betrouwbaarheid onvoldoende is, dan heeft het onderzoek minder of geen waarde.

#### **Beoordelingscriteria voor cohortonderzoek en patiënt-controleonderzoek**

Daar de meeste aspecten voor het beoordelen van cohort- en patiënt-controleonderzoeken bij beide typen ontwerpen hetzelfde zijn, worden ze hier gezamenlijk besproken, waarbij wordt aangegeven voor welk onderzoeksontwerp het geldt. Ook hier geldt dat als aan al de te beoordelen punten wordt voldaan, dit ten goede komt aan de validiteit. Wanneer niet aan de voorwaarden wordt voldaan, vermindert de interne validiteit van het onderzoek.

#### **Vergelijkbaarheid van de groepen**

Bij zowel cohortonderzoek als patiënt-controleonderzoek moeten de groepen zoveel mogelijk vergelijkbaar zijn. Uiteraard kunnen ze niet helemaal gelijk zijn, dat kan immers alleen als de onderzoeker ingrijpt en door middel van at random toewijzing bepaalt wie de onderzoeksvariabele heeft en wie niet. Echter, de controlegroep zou zoveel mogelijk moeten zijn samengesteld uit mensen van dezelfde populatie, met dezelfde achtergrond en uit patiënten in hetzelfde stadium van de ziekte. De groepen moeten daarom bij voorkeur niet te heterogeen van samenstelling zijn, zodat de éne patiënt niet in een veel verder stadium van de ziekte is dan de andere, of een andere vorm van de ziekte heeft.

#### **Gelijke metingen**

De bepaling van de uitkomstmaat moet in een cohortonderzoek onafhankelijk gebeuren van de groepssamenstelling. Dat wil zeggen dat degene die de metingen verricht, idealiter niet weet in welke groep de patiënt zit die hij of zij beoordeelt. Als dat niet mogelijk is, zou een goed alternatief zijn dat de beoordelaar niet op de hoogte is van de vraagstelling en zeker niet van de hypothese van het onderzoek.

In een patiënt-controleonderzoek kan de in de bijdrage van Aartman en Van Loveren (2007) al vermelde 'recall bias' een ongewenste vertekening opleveren. Dat wil zeg-

Onderzoeksontwerp	Scoringsmogelijkheden*			
<b>Gerandomiseerd gecontroleerd onderzoek</b>				
At random toewijzing	ja	nee	?	nvt
Gelijkheid van de groepen	ja	nee	?	nvt
Eenduidige interventie	ja	nee	?	nvt
Uitval	ja	nee	?	nvt
Blinding van de patiënt	ja	nee	?	nvt
Blinding van de effectbeoordelaar	ja	nee	?	nvt
Blinding van de behandelaar	ja	nee	?	nvt
Kwaliteit meetmethoden				
Betrouwbaarheid	voldoende	matig	onvoldoende	
Validiteit	voldoende	matig	onvoldoende	
Externe validiteit	voldoende	matig	onvoldoende	
Toepasbaarheid	voldoende	matig	onvoldoende	
<b>Cohortonderzoek en patiënt-controleonderzoek</b>				
Vergelijkbaarheid van groepen	ja	nee	?	nvt
Gelijke metingen	ja	nee	?	nvt
Duur en volledigheid follow-up	ja	nee	?	nvt
Kwaliteit meetmethoden				
Betrouwbaarheid	voldoende	matig	onvoldoende	
Validiteit	voldoende	matig	onvoldoende	
Externe validiteit	voldoende	matig	onvoldoende	
Toepasbaarheid	voldoende	matig	onvoldoende	
<b>Systematisch literatuuronderzoek</b>				
Heldere vraagstelling	ja	nee	?	
Zoekactie en selectie artikelen	voldoende	matig	onvoldoende	
Beoordeling methodologische kwaliteit	voldoende	matig	onvoldoende	
Onafhankelijke beoordeling	voldoende	matig	onvoldoende	
Combineren van resultaten	voldoende	matig	onvoldoende	

\*Ja: voldoende nauwkeurig beschreven; nee: niet beschreven of uitgevoerd, maar was wel mogelijk geweest; ?: niet te beoordelen, geen informatie beschikbaar; nvt: niet van toepassing, niet mogelijk.

Tabel 1. Beoordelingscriteria onderzoeksontwerpen.

gen dat patiënten zich van belang zijnde variabelen beter of anders kunnen herinneren dan controlepersonen. Daardoor ontstaat er een ongewenst verschil tussen de groepen. Ook kan het zijn dat de gegevens van wat later patiënten blijken te zijn, anders zijn genoteerd en dus anders worden gemeten dan van controlepersonen.

### Duur en volledigheid follow-up

In cohortonderzoek moet de duur van het onderzoek lang genoeg zijn om de ontwikkeling van de gekozen uitkomstmaat te kunnen bepalen (bijvoorbeeld cariës). Verder moet er, net als bij een gerandomiseerd gecontroleerd onderzoek, geen sprake zijn van te veel en/of selectieve uitval. Het risico van uitval bij een langdurig cohortonderzoek is groot. Indien dan ook nog eens sprake is van meer uitval van patiënten in de éne groep dan in de andere, waarbij dan niet is uit te sluiten dat het type behandeling daar verantwoordelijk voor is, is de waarde van het onderzoek op z'n zachts gezegd aan de magere kant.

### Kwaliteit meetmethoden

De beoordeling hiervan kan op dezelfde manier plaatsvinden als bij een gerandomiseerd gecontroleerd onderzoek.

### Beoordelingscriteria voor systematische literatuuronderzoeken

Een systematisch literatuuronderzoek van gerandomiseerde gecontroleerde onderzoeken staat het hoogste op de ladder van evidence. Omdat ook de kwaliteit van een systematisch literatuuronderzoek kan variëren, worden hier enkele aspecten genoemd op grond waarvan deze kan worden beoordeeld (zie ook Offringa et al, 2000). Als aan alle beoordelingscriteria is voldaan, zou het erop neer moeten komen dat het systematische literatuuronderzoek objectief is uitgevoerd en kan worden gerepliceerd, met dezelfde conclusie als gevolg.

### Heldere vraagstelling

Een goed systematisch literatuuronderzoek gaat uit van een expliciete vraagstelling die eenduidig te beantwoorden

is. Dat betekent dat duidelijk moet zijn in welke populatie welke interventie is bestudeerd en waar die mee wordt vergeleken. Tevens zou het moeten gaan om duidelijke uitkomstmaten, en niet om een algemeen begrip als effect van de behandeling.

### Zoekactie en selectie artikelen

Een systematisch literatuuronderzoek en een meta-analyse laten duidelijk de in- en exclusiecriteria zien die zijn gebruikt bij het selecteren van de artikelen. Het gaat hierbij om de populaties, de interventies, de onderzoeksontwerpen en de uitkomstmaten. Het is belangrijk dat alle onderzoeken in het systematische literatuuronderzoek en meta-analyse zijn meegenomen. De zoekstrategie moet dan ook gestructureerd zijn uitgevoerd en duidelijk vermeld worden in het overzicht, opdat hij kan worden nagedaan. De zoekactie moet op z'n minst gebruikmaken van een elektronische database als MEDLINE (PubMed).

Het is belangrijk dat van elk onderzoek dat bij de zoekstrategie wordt gevonden verantwoording wordt afgelegd waarom het onderzoek wel of niet is opgenomen. De lezer kan dan beoordelen of de juiste onderzoeken in het systematische literatuuronderzoek of meta-analyse zijn betrokken.

### Beoordeling methodologische kwaliteit

Van de onderzoeken die worden gebruikt moet een duidelijke analyse van de kwaliteit worden gedaan. Hierbij kan voor de primaire onderzoeksartikelen gebruik worden gemaakt van beoordelingslijsten voor de verschillende onderzoeksontwerpen, gelijkend op de lijst getoond in tabel 1. Eventueel vindt er een weging plaats van de criteria. Een dergelijk schema zou in elk systematisch literatuuronderzoek en meta-analyse terug te vinden moeten zijn.

### Onafhankelijke beoordeling

Idealiter zouden de zoekactie en de selectie van de artikelen door ten minste 2 onderzoekers onafhankelijk van elkaar moeten worden uitgevoerd. Ook de beoordeling van de methodologische kwaliteit zou door ten minste 2 onafhankelijke onderzoekers moeten worden verricht, waarbij ook gegevens over de interbeoordelaarsbetrouwbaarheid worden gerapporteerd.

### Combineren van resultaten

Het spreekt vanzelf dat wanneer onderzoeken gecombineerd worden, dit verantwoord en gewogen moet gebeuren, waarbij duidelijk wordt beschreven hoe dit is gebeurd. Gegevens van verschillende onderzoeken kunnen niet bij elkaar worden genomen wanneer de uitkomstmaat niet voldoende overeen komt, of de onderzoeken qua opzet te veel van elkaar verschillen. Ten slotte, mochten verschillende onderzoeken uiteenlopende resultaten geven, dan dienen het systematische literatuuronderzoek en de meta-analyse hier een duidelijke verklaring voor te geven.

### Externe validiteit

Het begrip externe validiteit moet niet verward worden met het begrip interne validiteit. Externe validiteit betreft de vraag of de resultaten van het onderzoek ook gelden buiten de omstandigheden van het onderzoek zelf. Indien de omstandigheden waaronder het onderzoek is uitgevoerd dermate specifiek zijn dat dit niet het geval is, is er sprake van een lage externe validiteit. Naarmate het meer aannemelijk is dat de resultaten ook van toepassing zijn buiten de gebruikte steekproef en omstandigheden, in andere gebieden, onder iets minder gecontroleerde omstandigheden, voor bredere leeftijdsgroepen enzovoorts, is de externe validiteit groter. Deze externe validiteit is verweven met het laatste aspect dat besproken gaat worden, de toepasbaarheid van het beschreven onderzoek voor de PICO-vraag.

### Toepasbaarheid voor de PICO-vraag

Naast de beoordeling van de validiteit van onderzoek, dient een beoordeling over de toepasbaarheid plaats te vinden. Afhankelijk van de casus waarop de PICO-vraag is gebaseerd, is een onderzoek meer of minder relevant voor de beantwoording van de PICO-vraag. Als in een onderzoek een steekproef wordt gebruikt van een type patiënt waar een tandarts-algemeen practicus nooit mee te maken heeft, is zo'n onderzoek niet relevant genoeg om te gebruiken. Verder heeft het ook geen zin een artikel te lezen waarin behandelmethoden worden gebruikt die in een tandartspraktijk niet voorhanden zijn of die een tandarts-algemeen practicus niet beheerst. Onderzoek dat niet klinisch is, is doorgaans ook niet voldoende toepasbaar, omdat de uitkomstmaten die in een dergelijk onderzoek worden gebruikt vaak surrogaatuitkomsten zijn. Dat wil zeggen dat de variabelen die worden gemeten wel een verondersteld verband hebben met een klinisch relevante uitkomstmaat, maar het effect ervan kan toch anders uitpakken dan in de praktijksituatie (bijvoorbeeld lekkage van een behandeld wortelkanaal in een onderzoek met geëxtraheerde gebitselementen versus de overleving van een endodontisch behandeld gebitselement in een *in vivo*-onderzoek).

Andere aspecten die de toepasbaarheid betreffen zijn de mogelijke bijwerkingen van producten, de voorkeuren en wensen van patiënten en de waardering van de patiënt voor de verschillende mogelijke uitkomstmaten. Dit zijn allemaal zaken die mede bepalend zijn voor het nemen van een beslissing omtrent de uiteindelijke behandeling of voorlichting, of voor het geven van het antwoord op de PICO-vraag.

Als de tandarts-algemeen practicus een artikel leest, is er een aantal specifiekere vragen dat hij kan beantwoorden om te bepalen of het uiteindelijk toch relevant genoeg is om te gebruiken bij het geven van het antwoord op de PICO-vraag. Een uitgebreidere beschrijving van de beoordelingspunten betreffende de toepasbaarheid voor onderzoeksartikelen uit elk domein is te vinden in Offringa et al (2000).



## Tot slot

In dit artikel is getracht de tandarts-algemeen practicus verder te helpen bij het kritisch lezen en beoordelen van wetenschappelijke tandheelkundige artikelen in het kader van het nemen van de juiste beslissing ten aanzien van behandeling van of advies aan patiënten. De belangrijkste boodschap daarbij is dat van een artikel vooral de sectie materiaal en methode relevante informatie geeft over de geldigheid van de resultaten en conclusies die door de auteurs van artikelen worden gepresenteerd. Alleen als het onderzoek op juiste waarde wordt geschat door de opzet goed te begrijpen en te beoordelen, kan worden bepaald of het betreffende onderzoek werkelijk een antwoord kan geven op de gestelde PICO-vraag.

## Literatuur

- > Aartman IHA, Loveren C van. Onderzoeksonwerpen en de ladder van evidence. Ned Tijdschr Tandheelkd 2007; 114: 161-165.
- > Forrest JL, Miller SA. Evidence-based decision making in action: Part 2 - Evaluating and applying the clinical evidence. J Contemp Dent Pract 2003; 4: 42-52.
- > Hackshaw AK, Paul EA, Davenport ES. Evidence-based dentistry. An introduction. Oxford: Blackwell Munksgaard, 2006.
- > Hunt RJ. Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability. J Dent Res 1986; 65: 128-130.
- > Kuiper C, Verhoef J, Louw D de, Cox K. Evidence-based practice voor paramedici. Methodiek en implementatie. Utrecht: Lemma, 2004.
- > Mileman PA, Hout WB van den. Evidence-based diagnostiek en klinische besluitvorming. Ned Tijdschr Tandheelkd 2007; 114: 187-194.
- > Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Clin Oral Investig 2003; 7: 2-7.
- > Offringa M, Assendelft WJJ, Scholten RJPM. Inleiding in evidence-based medicine. Klinisch handelen gebaseerd op bewijsmateriaal. Houten: Bohn Stafleu Van Loghum, 2000.

## Summary

### Evidence-based dentistry and rating the validity and applicability of research

In order for a general dental practitioner to give a valid answer to a question asked according to the Problem-Intervention-Comparison-Outcome (PICO)-system, three questions need to be asked regarding validity, relevance and applicability of studies. The validity is determined by the position of the research on the ladder of evidence, and also by the assessment of the methodological quality of the study at stake. In this assessment, the study is rated on the basis of specific criteria. A randomized controlled trial for example can be assessed with regard to the randomisation procedure, blinding of patients, raters and treatment, and the number of patients lost for follow-up. In cohort studies and case-control studies the criteria on the basis of which an article can be assessed are the comparability of groups, the degree of independence among measurements, and the length and thoroughness of the follow-up. Furthermore, in all studies it is important that the quality of the measurement instruments is good. The assessment criteria for systematic reviews are a clear research question, a clearly written and repeatable search action for articles and an independent assessment of the methodological quality of the included articles. Finally, the general dental practitioner in determining applicability needs to pay attention to the sample and the type of patient used in the research.

## Bron

I.H.A. Aartman<sup>1</sup>, C. van Loveren<sup>2,3</sup>

Uit <sup>1</sup>de afdeling Sociale Tandheelkunde en Voorlichtingskunde en <sup>2</sup>de afdeling Cariologie Endodontologie Pedodontologie van het Academisch Centrum Tandheelkunde Amsterdam (ACTA) en <sup>3</sup>het Ivoren Kruis

Datum van acceptatie: 12 februari 2007

Adres: mw. dr. I.H.A. Aartman, ACTA, Louwesweg 1, 1066 EA Amsterdam  
i.aartman@acta.nl