

Clinical Outcomes Studies voor CAM

Opzet en verslaglegging

In vergelijking met dubbelblind studies vormen Clinical Outcomes Studies (COS) in veel CAM-onderzoek een meer aangewezen methode voor evaluatie van effectiviteit.

C.W. Kramers

Inleiding

Hoe toets je het effect van alternatieve behandelmethoden? 'Outcomes research' zonder controlegroepen verdient hier vooralsnog de voorkeur boven de traditionele 'randomized clinical trial'. Waar moet de potentiële onderzoeker zoal aan denken bij de opzet en uitvoering van een 'Clinical Outcomes Study'? Voor de beantwoording van deze vraag verplaatsen we ons in de positie van de onderzoeker als hij het afsluitende onderzoeksverslag moet schrijven. Daarbij komt ook aan bod hoe het onderzoek wordt opgezet en hoe de gevonden resultaten op schrift worden gesteld.

Het meten van de werkzaamheid van medische behandelingen, of dat nu regulier-medische of CAM-behandelingen zijn, verloopt in principe op een en dezelfde wijze (de letters 'CAM' staan voor 'Complementary and Alternative Medicine'). Bij elke werkzaamheidsstudie in de geneeskunde draait het immers altijd om het vergelijken van de klachten voor en na een behandeling (de experimentele interventie) bij een vooraf gedefinieerde groep patiënten.

Een dergelijke studie kan, afhankelijk van de vraagstelling, echter op verschillende manieren plaatsvinden. Soms is een rigoureuze onderzoek nodig met controlegroepen, randomisatie en blinding; maar men kan ook volstaan met een eenvoudiger, meer praktijkgerichte observationele onderzoeksopzet, zonder controlegroep. De manier waarop het onderzoek uiteindelijk plaats vindt – wel of niet gecontro-

leerd – wordt bepaald door een veelheid aan factoren. In de eerste plaats zal de gebruikte onderzoeksmethode afhangen van de vraagstelling van het onderzoek. Verder spelen allerlei maatschappelijke en culturele factoren een rol bij dit keuzeprocess, zoals ethische overwegingen of de beschikbare financiën.

De 'Clinical Outcomes Study' (COS) en de 'Randomized Controlled Trial' (RCT)

Voor het toetsen van CAM-therapieën lijkt ons de weg van de observationele, praktijkgerichte, ongecontroleerde 'outcomes research' een goede keuze. Een concreet onderzoek binnen de 'outcomes research' wordt wel 'Clinical Outcomes Study' (COS) genoemd.

Het is beslist niet nodig altijd gebruik te maken van de zogenaamde gouden standaard onder de onderzoeksmethoden, te weten: de 'randomized controlled trial' (RCT). Voor een gebied als CAM waar nog nauwelijks adequate observationele studies werden verricht, is de COS een goede keuze.

In het algemeen beschouwt men 'outcomes research' als voorportaal voor meer geavanceerd onderzoek. Via zulk vooronderzoek tracht men dan categorieën patiënten te identificeren die gunstig op de behandeling reageren, waarna men dan besluit wel of geen gecontroleerde studie te doen. Ook is het mogelijk via een COS een idee te krijgen van het werkingsmechanisme van een bepaalde behandeling. Natuurlijk kleven er wel beperkingen

aan een COS. Zonder controlematregelen kan men nu eenmaal geen vergelijking maken met een bestaande of een placebobehandeling, waarmee men de diverse vormen van bias (vertekening van de resultaten) beoogt uit te sluiten. Toch kan een COS, zoals hierboven betoogd, waardevolle gegevens opleveren, ook als er geen 'randomized clinical trial' op volgt. Daarnaast heeft 'outcomes research' enkele interessante voordelen boven de traditionele RCT-methodologie. Zo kan in tegenstelling tot een RCT een COS in het algemeen relatief gemakkelijk in de dagelijkse praktijk van behandelaars worden ingebed. Ook zijn de kosten van een ongecontroleerde studie meestal flink lager. Daarnaast hoeft de onderzoeker zich niet druk te maken over een aantal methodologische hoogstandjes, zoals matchen, randomiseren en het dubbel blinderen van patiënten en behandelaars, met alle ethische implicaties van dien. Trouwens dat blinderen van behandelaars druist o.i. in tegen alle beginselen van de geneeskunst. Van de behandelaar mag immers worden verwacht dat hij, niet alleen in de praktijk maar juist ook in onderzoekssituaties, bij wijze van spreken zijn ogen wijd open houdt voor wat er allemaal met de aan hem toevertrouwde patiënt gebeurt.

Een ander voordeel van 'outcomes research' ten opzichte van onderzoek met controlegroepen is dat proefpersonen langdurig gevolgd kunnen worden, waardoor ook op de lange termijn de effecten van een therapie gemeten kunnen worden. Dit is vooral van belang als de interventie gepaard gaat met een reeks van regelmatige evaluatie- en bijstuurmomenten. Als de onderzoeker in staat wordt gesteld een dergelijk compleet proces mee te beleven, kan bij hem en bij de patiënt een echt gevoel van evidentie ontstaan ten aanzien van het resultaat. Een dergelijke situatie is moeilijk denkbaar in het geval van een dubbelblinde RCT. Hier kan men het onderzoek immers niet te lang laten duren juist vanwege de genoemde hoogstandjes: er bestaat een toenemend risico op 'lekken' in de blinding van de proefpersonen naarmate de studie langer duurt. Verder is de ongecontroleerde studie geknipt als in-

stapmodel voor de onervaren CAM-onderzoeker. De fouten die het moeilijke vak van onderzoeker nu eenmaal met zich meebrengt, hebben in een COS-setting meestal minder vervelende (financiële) consequenties. Dit punt van het instapmodel is van groot belang omdat er dringend behoefte is aan de vorming van een wetenschappelijke traditie en infrastructuur op het gebied van CAM-werkzaamheidstoetsen.

Het is dan ook te betreuren dat 'outcomes research' vanwege haar ongecontroleerde karakter nogal eens wordt afgedaan als een ondoelmatige onderzoeksmethode. Deze methodekritiek lijkt echter eerder van 'religieuze' dan van onderzoekstechnische aard te zijn. Nog steeds 'gelooft' een substantieel deel van het huidige wetenschappelijke establishment alleen in zogenaamde evidence-based medicine (EBM). Voor de goede orde: er is niets tegen EBM, hoe meer evidentie hoe beter. Maar er is wél wat tegen EBM als men het begrip vernauwt tot een geneeskunde die uitsluitend gebaseerd is op de resultaten van gerandomiseerde gecontroleerde studies met al hun praktische, methodologische en ethische bezwaren. EBM is meer dan dat. Het is in de woorden van Sackett et al (1996) een stoel met drie poten: 'patient choice', 'clinical judgement' en 'best available evidence'. En met 'best available evidence' doelt men dan op de uitkomsten van RCT's. Onzes inziens kan en mag de poot van de 'neutrale' bewijzen van gerandomiseerde studies nooit de andere twee poten – de persoonlijke wensen van de patiënt en het persoonlijke oordeel van de behandelaar – overrulen. Deze beide laatste factoren kunnen wél heel goed ondersteuning vinden via 'outcomes research'. Gelukkig zien langzamerhand steeds meer wetenschappers de COS als een nuttig instrument bij de bepaling van de evidence-base. Zo acht men 'outcomes research' met name geïndiceerd als de verwachte behandel-effecten (zeer) groot zijn. Een gecontroleerde studie is dan niet nodig omdat de effecten wel duidelijk zijn. Zo waren destijds de effecten van penicilline zo onmiskenbaar dat nauwkeuri-

ge observatie genoeg was. Een RCT is ook niet nodig als het natuurlijk beloop van een ziekte bekend is en een goede schatting van de prognose is te maken of als de baseline redelijk stabiel is. In die gevallen kunnen duidelijke en aanhoudende gunstige effecten, mits goed gedocumenteerd gedurende een langere observatieperiode, zeer overtuigend zijn (Walach et al, 2002).

De ‘Clinical Outcomes Study’: opzet en rapportage

We behandelen in deze bijdrage een reeks aandachtspunten die van belang zijn bij het opzetten van een ‘Clinical Outcomes Study’ en dan speciaal gericht op de werkzaamheid van CAM-therapieën. Nu is het gebruikelijk om bij de voorbereiding van een onderzoek de verschillende aandachtspunten langs te lopen op geleide van een checklist voor het maken van onderzoeksvoorstellen (onderzoeksprotocollen). Voor ons betoog wijken we van dit gebruik af en nemen we als raamwerk het onderzoeksverslag zoals dat aan het eind van een studie geschreven wordt. We geven hier de voorkeur aan de afsluitende rapportage boven het onderzoeksprotocol bij de start van een studie omdat we ook expliciet aandacht willen besteden aan het produceren van een dergelijke rapportage; en de reden daarvoor is dat ‘alternatieve’ onderzoekers niet altijd even consciëntieus te werk gaan bij het schrijfproces. Menig verslag laat qua vorm en inhoud te wensen over. Dat geldt trouwens niet alleen voor ‘alternatieve’ evaluatie-onderzoeken. Ook veel ‘reguliere’ rapportages bevatten vaagheden en fouten. Zo zijn tabellen en grafieken voor de lezer vaak moeilijk te ontcijferen. Soms is uit het verslag niet eens te achterhalen hoe het onderzoek precies in zijn werk ging. Het is dan ook, zeker voor een beginnende onderzoeker, raadzaam zich niet alleen op methodologisch en statistisch, maar vooral ook op schrijftechnisch gebied te laten bijstaan. Een uitspraak van Paton (1979) spreekt hier boekdelen: “doing the research [...] is child’s play compared with the moment of truth when you come to write

up [...] your results.”

Om het geheel nog meer te concretiseren, nemen we als raamwerk *niet* het gebruikelijke en vaak uitvoerige onderzoeksrapport, maar het afsluitende tijdschriftartikel, waarin de onderzoeker, vanwege de beperkte ruimte, nog meer tot de kern moet komen en zich zo exact en bondig mogelijk dient uit te drukken.

Op basis van langjarige ervaring met het begeleiden van ‘alternatieve’ onderzoekers adviseren wij diegenen die nog weinig of geen expertise in het doen onderzoek hebben kunnen opbouwen, al vast op voorhand het onderdeel ‘resultaten’ van de eindrapportage te schrijven, inclusief tabellen en figuren met verwachte of fake-data. Dit draagt er ongetwijfeld toe bij dat de verschillende analyses correct gebeuren zonder “torturing the data until it confesses” (Vickers, 2002). Daarnaast vragen we de potentiële onderzoeker meestal ook alvast een concept-samenvatting te schrijven. Dat is voor menig een even wennen maar het levert soms zeer verrassende ‘ontdekkingen’ op. Zodanige ontdekkingen zelfs, dat het hele onderzoek zoals men dat in gedachten had vaak compleet op de helling moet. In het algemeen bestaat een wetenschappelijk artikel waarin de resultaten van een klinische studie worden gepresenteerd uit vijf onderdelen, te weten: samenvatting, inleiding, methoden, resultaten, discussie. (Met daarnaast zaken zoals titel, auteursgegevens, literatuurlijst.) Deze onderdelen zijn dan nog weer uitgesplitst in aandachtspunten (zie op p.238 het kader ‘Outline voor een tijdschriftartikel’). Per onderdeel zullen de verschillende aandachtspunten bij de rapportage van een COS hieronder de revue passeren. In de tekst voorzagen we de verschillende aandachtspunten van aparte tussenkopjes, die in het definitieve artikel ook heel goed kunnen worden weggelaten of vervangen door andere kopjes. Ze dienen hier alleen voor de systematiek in ons betoog.

Outline voor een tijdschriftartikel met de resultaten van een 'Clinical Outcomes Study'

basisgegevens

- titel onderzoek, 'key-words', namen auteurs, correspondentie-adres, dankwoord, etc.

samenvatting

- inclusief 'summary'

inleiding —*why did you start?*

- aanleiding
- doel en vraagstelling
- ziektebeeld (symptomen, oorzaken, verspreiding, etc.)
- behandel mogelijkheden (regulier, alternatief)
- de onderzochte (getoetste) behandeling

methoden —*what did you do?*

- setting en onderzoekers
- de proefpersonen (selectie, rekrutering, informed consent)
- design (controles, blinding, retrospectief, prospectief, pilot, etc.)
- experimentele interventie (getoetste behandeling)
- uitkomstmaten (variabelen)
- dataverwerving (meetinstrumenten, meetmomenten, meetschalen, steekproeven)
- dataverwerking (statistische analyse)

resultaten —*what did you find?*

- prognostische patiëntgegevens
- behandelgegevens
- 'response-data'

discussie —*what does it mean?*

- conclusie
- beperkingen van het onderzoek
- mogelijk vervolgonderzoek

literatuur en evt. bijlagen

Eerste onderdeel van het artikel: de samenvatting

De samenvatting van het artikel bevat in kort bestek de belangrijkste feiten betreffende het verrichte onderzoek. Sommige tijdschriften stellen als eis dat de samenvatting dezelfde indeling heeft als die van het gehele artikel. Ook wordt meestal een Engelse 'summary' verlangd. Speciale aandacht moet worden besteed aan de conclusies van het onderzoek, omdat de lezer die nu eenmaal het eerst wil kennen. Zijn die conclusies onduidelijk geformuleerd, dan zullen lezers gauw afhaken. Let wel: het is niet erg als de studieresultaten slechts tot vage conclusies leiden, het is wel erg als daar in de tekst van het artikel en dus ook in de samenvatting op zichzelf vaag over gedaan wordt.

Tweede onderdeel van het artikel: de inleiding

Aanleiding

In dit onderdeel dient de auteur antwoord te geven op de vraag: 'why did you start?' De inleiding van een artikel waarin de uitkomsten van klinisch onderzoek worden gepubliceerd, begint meestal met een paar zinnen over de aanleiding tot dat onderzoek. Een dergelijke aanleiding kan zijn: positieve ervaringen met een bepaalde therapie of een vraag van derden (bijvoorbeeld een patiëntenvereniging), of de wens de resultaten van andere onderzoekers te bevestigen. Daarnaast is het schrijven van een thesis of dissertatie een belangrijke reden voor het doen van een wetenschappelijk onderzoek. In de praktijk zal de aanleiding voor een studie een combinatie van deze en nog andere factoren zijn.

Het spreekt vanzelf dat in het artikel ook ergens melding wordt gemaakt van eventuele opdrachtgevers resp. sponsors. Het is enigszins arbitrair op welke plaats dat gebeurt. Soms wordt in een aparte noot aan het eind van het artikel de naam van een sponsor vermeld. Als bedrijven apparatuur of medicijnen voor het onderzoek beschikbaar stellen, wordt dat meestal aangegeven in de para-

graaf 'methoden'.

Doel en vraagstelling

De belangrijkste elementen van de inleiding van het artikel zijn doel en vraagstelling van het verrichte onderzoek. Volgens de meeste auteurs is het doel van een COS het bevestigen van een klinische observatie. Men wil antwoord op de vraag of een duidelijke klachtenreductie bij een of meer patiënten na een bepaalde behandeling ook optreedt bij grotere aantallen patiënten om zo verdere *gecontroleerde* studies te rechtvaardigen. Anderzijds kunnen o.i. ook studies zonder controlegroep op zichzelf al voldoende 'evidence base' opleveren wanneer gebruik wordt gemaakt van adequate, liefst gevalideerde, meetmethoden en wanneer de resultaten nauwkeurig geregistreerd zijn en betrekking hebben op een flinke, goed gedefinieerde, studiepopulatie. In het concrete geval zal men als doel van welke klinische studie dan ook simpel aangeven dat men wil weten in hoeverre behandeling X werkzaam is bij patiënten met aandoening, symptoom of klacht Y. Daarna geeft men in de vraagstelling nog wat concreter aan om welke patiëntenpopulatie en om welke behandelmethode het in de studie precies ging. Ook wordt vermeld welke uitkomstmaten of variabelen (aantal huiluren, bloeddruk, 5-jaaroverleving, algemeen welbevinden, etc.) men gebruikte. De invulling van een concrete vraagstelling dwingt de onderzoeker tot het maken van een aantal keuzen die hij waarschijnlijk, al of niet door de omstandigheden gedwongen, reeds lang gemaakt heeft. Ook als men voor een ongecontroleerde studie heeft gekozen, doet de onderzoeker er goed aan voordat hij de definitieve opzet ontwerpt nog een aantal zaken te overwegen. Bepaalde keuzes hebben namelijk consequenties voor die opzet, bijvoorbeeld voor de setting. Zo kan een heel interessante onderzoeksvraag zijn, of de experimentele CAM-behandeling van nut kan zijn als ondersteuning (adjuvans) bij een bestaande (reguliere) therapie. Bij een dergelijke vraagstelling zal men op de een of andere manier moeten samenwerken met of in een regu-

liere setting. Maar de onderzoeker kan ook geïnteresseerd zijn in de vraag of een bepaalde 'alternatieve' *diagnostische* meetmethode (kinesiologie, pendel, medicamententest) reproduceerbare resultaten geeft. Men kan ook onderzoek doen naar de kosteneffectiviteit of naar mogelijke ongewenste bijwerkingen van een behandeling, bijvoorbeeld op een gebied als de fytotherapie, waar de laatste tijd veel commotie heerst over (vermeende) toxische effecten en interacties met reguliere middelen. Nog een heel andere keuze in verband met de onderzoeksvraag is of men zich beperkt tot een studie in één praktijk met misschien een of twee behandelaren, of dat er meer centra met meer behandelaren bij betrokken worden. Wat het punt van de bijwerkingen betreft vragen wij ons trouwens af waarom het (reguliere en complementaire) klinische toetsingsonderzoek zich voortaan niet zou kunnen beperken tot het nauwgezet opsporen en identificeren van ongewenste en schadelijke effecten van behandelmethoden; de werkzaamheid of onwerkzaamheid blijkt immers wel uit de praktijk. In het kader van dit artikel kunnen wij helaas verder niet op dit punt ingaan.

Voordat men met het 'echte' onderzoek begint kan het soms nuttig zijn eerst eens een pilot- of 'feasibility'-studie te doen bij een klein groepje proefpersonen. Dergelijke studies hebben meestal een specifiek 'technisch' doel. Zo kan men ervaring willen opdoen bijvoorbeeld met het (mondeling of schriftelijk) informeren van patiënten of met een bepaalde meetmethode (bijvoorbeeld het afnemen van een vragenlijst) of met het registreren van data (bijvoorbeeld het verwerken en analyseren van antwoorden op een vragenlijst). Een pilotstudie kan ook zeer verhelderend zijn als met meer behandelaren wordt gewerkt, vooral als die er onderling verschillende therapeutische strategieën op na blijken te houden. Dat is trouwens wel een lastig punt in de CAM-wereld, waar het soms compleet onduidelijk is hoe een behandelaar precies te werk gaat. Die werkwijze zal, wil een wetenschappelijk onderzoek zin hebben, eerst zo goed mogelijk geëxpliciteerd moeten

worden in een aparte werkwijze-studie waarin soms uitvoerige gesprekken met behandelaars nodig zijn. Nadat deze en andere keuzen (impliciet dan wel expliciet) zijn gemaakt, zal de onderzoeker in de uiteindelijke vraagstelling tenslotte duidelijk aangeven welke onderzoeksuitkomsten als positief of gunstig moeten worden beschouwd en welke niet. Dit begrip ‘gunstig’ heeft twee dimensies: een voor de patiënten gunstig onderzoeksresultaat is pas werkelijk positief als dat resultaat niet op toeval berust (1) en tegelijkertijd klinisch relevant is (2). Of een gunstige (of ongunstige) uitkomst wel of geen toevalsbevinding is, bepalen we in het huidige statistische tijdsgewricht met een significantietoets. Is de onderzoeker van plan een dergelijke toets uit te voeren, dan kan de vraagstelling van het onderzoek geformuleerd worden in de vorm van twee elkaar uitsluitende hypothesen: een nulhypothese en een alternatieve hypothese. De nulhypothese gaat ervan uit dat de behandeling, statistisch gezien, niet het beoogde effect heeft. De alternatieve gaat juist wel van een positief resultaat uit. Na afloop van het onderzoek wordt een van beide hypothesen verworpen en de andere aanvaard.

Wat de tweede dimensie van het begrip ‘gunstige uitkomst’ betreft, de klinische relevantie: dat is een heel ander verhaal. Mocht een significantietoets leiden tot verwerping van de nulhypothese en aanvaarding van de alternatieve hypothese – een positieve uitkomst is statistisch gezien reëel en berust niet op toeval – dan is het heel goed mogelijk dat het positieve resultaat van de studie voor de praktijk weinig betekenis heeft. Dit soort situaties zien we bijvoorbeeld bij zeer grote studiepopulaties. Wanneer de gezondheid van de patiënten na de behandeling gemiddeld maar heel weinig (klinisch niet relevant) verbeterd is, kan een dergelijk relatief klein verschil tussen vóór en ná de behandeling toch statistisch significant zijn, door het grote aantal patiënten. Het zal duidelijk zijn dat de onderzoeker in een dergelijk geval niet tevreden is met het resultaat. Die klinische relevantie van studie-uitkomsten is overigens lang niet altijd even gemakke-

lijk te bepalen. Het is vooral lastig als er geen resultaten van eerder vergelijkbaar onderzoek beschikbaar zijn zoals we vaak zien op het gebied van CAM-behandelingen. In dat geval kan de onderzoeker niet terugvallen op de literatuur en zal hij, zoveel mogelijk in overleg met (andere) behandelaars en onderzoekers – en niet te vergeten de patiënten zelf –, van te voren moeten bepalen welke verbetering als realistisch en relevant beschouwd kan worden.

Hieronder geven we een voorbeeld van een vraagstelling waarbij bovengenoemde aspecten aan bod komen. De vraagstelling werd ontleend aan een lopende studie naar het effect van een osteopathische behandeling bij huilbaby's (Batstra, 2004).

- nulhypothese (H₀): een osteopathische behandeling bij overmatig huilende baby's leidt niet tot reductie van het aantal huiluren
- alternatieve hypothese (H₁): een osteopathische behandeling bij overmatig huilende baby's leidt tot reductie van het aantal huiluren
- op basis van de beschikbare literatuur en eigen praktijkervaringen beschouwen wij het resultaat van een osteopathische behandeling bij de voor dit onderzoek geselecteerde populatie huilbaby's klinisch relevant, indien het aantal huiluren na drie behandelingsessies bij 50% van de huilbaby's tot de helft kan worden teruggebracht

Ziektebeeld

Wanneer het onderzoek zich richt op een bepaalde ziekte, aandoening, klacht of symptoom, wordt van de auteur van het onderzoeksverslag verwacht dat hij een beschrijving geeft van het betreffende ziektebeeld. Hij dient dat zo uitgebreid mogelijk te doen. Gaat het bijvoorbeeld om het effect van een bepaalde therapie bij migraine of bij een meer acute zaak als huilbaby's, dan zal men eerst een uitvoerige toelichting geven bij de ziekte migraine resp. het fenomeen ‘huilbaby’. Om te beginnen geeft men de officiële definitie, als die er is. In het boven gegeven voorbeeld van de huilbaby: het kind huilt per dag meer dan drie uur constant; meer dan

drie dagen per week; langer dan drie weken achtereen. Ook voor de diverse vormen van migraine bestaan dergelijke algemeen aanvaarde definities. Vervolgens zal men iets moeten meedelen over de oorzaken van de aandoening. En als die oorzaak niet bekend is of omstreden of meervoudig (multicausaal), zoals bij migraine, is het raadzaam een dergelijk feit ook in het verslag te vermelden. Men moet *juist* als CAM-onderzoeker beseffen dat het doen van onderzoek altijd is ingebed in een medisch-wetenschappelijke ontwikkeling. Vandaar dat het noodzakelijk is uitgebreid stil te staan bij de 'state of the art', welke feiten al bekend zijn over de ziekte en of er wellicht discussie over die feiten is. Zo worden in het geval van migraine recent vraagtekens geplaatst bij de vigerende opvatting dat de oorzaak gezocht moet worden in een vaatverwijding in de hersenen; een nieuwe theorie stelt dat die vaatverwijding geen oorzaak is maar eerder een gevolg van een andere etiologische (mogelijke neurologische) factor. Ook over de oorzaak van overmatig huilen bij baby's is discussie. In de medische wereld is het gebruikelijk het overmatig huilen te wijten aan darmkrampen ('infantile colic'). Anderen zien als oorzaak echter oververmoeidheid bij het kind door allerlei oorzaken, bijvoorbeeld doordat het kind teveel blootstaat aan prikkels vanuit de omgeving. Ook zou het fenomeen huilbaby in bepaalde gevallen op een geboortetrauma wijzen. Als de aard en de mogelijke oorzaken van het ziektebeeld op deze wijze voldoende in kaart zijn gebracht, kan worden ingegaan op de omvang van het probleem: hoe vaak en onder welke omstandigheden het voorkomt, hoe ernstig de symptomen worden ervaren en wat patiënten of ouders er gewoonlijk aan doen.

Behandel mogelijkheden

Na een beschrijving van het ziektebeeld zal de onderzoeker in zijn inleiding stilstaan bij de verschillende behandelingsmogelijkheden voor de aandoening. Men zal daarbij het hele arsenaal aan therapeutische richtingen de revue laten passeren,

zowel reguliere als complementaire methoden, met ook (indien relevant) aandacht voor psychotherapeutische en paramedische benaderingen, zoals fysiotherapie en logopedie.

Er wordt uitvoerig melding gemaakt van hetgeen het wetenschappelijk onderzoek naar resultaten van deze behandelingen tot nu toe heeft opgeleverd. Van belang is te vermelden welke literatuurbronnen (databanken) men daartoe geraadpleegd heeft en welke eventueel niet.

De onderzochte (getoetste) behandeling

Vanzelfsprekend komt in de inleiding van het artikel ook de geteste behandeling aan bod. Hierboven bij de aanleiding tot de studie heeft de auteur van het artikel al vermeld waarom juist deze therapie onderzocht werd. Op deze plaats kan hij verder op die therapie ingaan. Wat houdt hij precies in? Waar, sinds wanneer (historische gegevens) en door wie wordt hij vooral toegepast en voor welke indicaties? Hoe denkt men in reguliere kringen over deze therapie? Zijn er doelgroepen die er vooral baat bij lijken te hebben? Zijn er bijwerkingen bekend? Is er wetenschappelijk onderzoek naar gedaan? Als uit de beschikbare literatuur blijkt dat met deze therapie bij deze aandoening (bijvoorbeeld osteopathie bij huilbaby's) nog geen onderzoek gedaan werd, dan doet men er goed aan dat ook expliciet in het verslag te vermelden.

Het kan natuurlijk zijn dat het in de studie die men gaat doen niet zozeer om een bepaalde (*regulier*) gedefinieerde klinische diagnose gaat, zoals migraine of huilbaby, maar om een heel anders (*alternatief*) geformuleerde beschrijving van het probleem van de patiënt. Dit speelt vooral bij CAM-behandelmethoden die er een eigen specifiek diagnostisch systeem op na houden, zoals bijvoorbeeld de Traditionele Chinese Geneeskunde, de homeopathie of de antroposofische geneeskunde. Bij deze vormen van geneeskunde is de behandelaar meestal niet primair gefocust op de actuele aandoening, maar vooral ook op heel andere kenmerken die min of meer rechtstreeks aansluiten bij het therapeu-

tisch arsenaal van die behandelaar. Zo wil de klassieke acupuncturist bijvoorbeeld weten of zijn patiënt overwegend een yin- of juist een yang-type is. En afhankelijk van deze en andere constitutionele gegevens wordt bepaald waar en hoe de naalden worden gezet. Ook homeopaten en antroposofische artsen inventariseren het geheel van constitutionele en vooral ook biografische kenmerken van de patiënt. En juist dit geheel van kenmerken, waar de actuele klachten ‘onderdeel’ van uitmaken, verwijst dan naar een passende behandeling. Soms ook laten zich diagnose en therapie tot in het gebruikte jargon niet of nauwelijks als aparte entiteiten onderscheiden, zoals bij de klassieke homeopathie: via anamnese, onderzoek en repertorisatie identificeert de homeopaat bij zijn patiënt uit de veelheid van geneesmiddelbeelden het meest passende, bijvoorbeeld het Thuja-beeld. En deze Thuja (levensboom) is dan ook meteen het homeopathicum dat in een adequate bereiding, dosering en potentie als eerste middel wordt ingezet.

Derde onderdeel van het artikel: de methoden

Nadat in de inleiding van ons artikel de vraag: ‘why did you start?’ werd beantwoord, geven we in de paragraaf ‘methoden’ antwoord op de vraag: ‘what did you do?’ De lezer wil precies weten hoe het onderzoek in zijn werk is gegaan en wil niet blijven zitten met vragen die niet beantwoord worden of waar vaag over gedaan wordt. Hij wil zo nauwkeurig geïnformeerd worden over wat er met de proefpersonen en hun gegevens is gebeurd gedurende het hele traject vanaf hun rekrutering tot aan de laatste vragenlijst die ze moesten invullen. Sommige auteurs en tijdschriftredacties plaatsen boven dit onderdeel het kopje: ‘patiënten en methoden’.

Setting en onderzoekers

Om te beginnen zal het onderzoeksverslag een beschrijving geven van de setting waarin de studie heeft plaats gevonden. In de officiële eerste lijn, of ging het misschien om een ziekenhuis- of verpleeg-

huispopulatie? Of betrof het een consultatieve praktijk? Wat voor praktijk is dat dan? Waar bevindt die praktijk zich? Welke therapieën worden daar gegeven en door wie? Wat is de status van de participerende behandelaar(s) resp. onderzoeker(s)? Is de behandelaar ook de onderzoeker? Het is om allerlei redenen overigens beter de rollen van behandelaar en onderzoeker te scheiden. De behandelaar werkt dan met de onderzoeker mee en legt de gegevens vast zoals dat onderling is overeen gekomen. De onderzoeker kan deze gegevens aanvullen met bijvoorbeeld laboratoriumwaarden en vragenlijstgegevens. Een heel andere vraag is of de behandelaar speciaal moet zijn opgeleid (gespecialiseerd)? Is hij representatief voor de beroepsgroep of niet? Is hij lid van een beroepsvereniging met een vorm van tuchtrecht en is daar nascholing verplicht? Ontving hij op enigerlei wijze een beloning voor zijn inspanningen? Waren proefpersonen en onderzoekers verzekerd? De auteur van het artikel kan op deze punten niet duidelijk genoeg zijn.

Proefpersonen

Van groot belang is dat de schrijver van het artikel duidelijk maakt hoe de patiënten bij het onderzoek werden betrokken. Welke voorselectie daarbij werd toegepast, hoe de proefpersonen gerekruteerd werden, hoe lang de rekruteringsperiode duurde en via welke in- en exclusiecriteria de uiteindelijke studiepopulatie tot stand kwam. In het geval van een therapie met zeer specifieke diagnostiek kan het soms zinvol zijn proefpersonen te selecteren via een dubbele selectieprocedure. Deze heeft tot doel een zo homogeen mogelijke studiepopulatie te creëren, afgestemd op de betreffende behandelmethode. De procedure bestaat uit twee stappen. In een eerste stap worden proefpersonen geselecteerd op grond van een reguliere diagnose, bijvoorbeeld astma. De zo ontstane groep proefpersonen wordt nu, alternatief gezien, ‘gehomogeniseerd’ doordat alleen die proefpersonen worden geïncludeerd die tevens beantwoorden aan een bepaald alternatief criterium, bijvoorbeeld de homeopathische ‘diagnose’ Thuja-

beeld. Het nadeel van deze procedure is dat de populaties erg klein kunnen worden (Wiegant et al, 2002). Een ander probleem is dat men licht het ver-wijft krijgt hetzelfde te doen als onderzoekers van gerandomiseerde studies. Die passen ook vormen van progressieve selectie van proefpersonen toe (strengere uitsluitcriteria). Hiermee beoogt men dan studiepopulaties te creëren die optimaal geschikt zijn gemaakt om binnen het model van een RCT positieve resultaten op te leveren.

Nu zal het voor de praktijk van een ongecontroleerde CAM-studie vaak zo zijn dat in principe alle in aanmerking komende patiënten die zich gedurende een bepaalde periode voor een bepaalde behandeling melden, in het onderzoek worden meegenomen. Als de studie duidelijk gericht is op een specifieke medische aandoening, zoals huilbaby of migraine, zal de (niet-medisch gekwalificeerde) onderzoeker erop moeten letten dat hij alleen proefpersonen meeneemt met een door een arts (evt. medisch specialist) gestelde diagnose. Andere criteria waaraan proefpersonen moeten voldoen, kunnen zijn: bepaalde leeftijdsgrenzen, geen andere ernstige ziektes, geen andere behandelingen, etc..

In het voorbeeld van de al eerder genoemde osteopathiestudie was sprake van één behandelaar-onderzoeker die in principe alle kinderen selecteerde die zich met de reguliere diagnose 'huilbaby' gedurende een periode van tien maanden bij de praktijk meldde in de leeftijd van drie weken tot drie maanden. Voor de uiteindelijke analyse werden alleen die baby's geïncludeerd die de drie geplande behandelsessies afmaakten en voor wie de ouders beide vragenlijsten invulden.

Het is overigens niet altijd noodzakelijk alleen 'prospectief' proefpersonen te werven. Men kan ze ook 'retrospectief' rekruteren als ze al eerder bij de dokter waren. Vooral voor personen met een stabiele chronische aandoening is dit een mogelijkheid. Wel moeten behandelaars dan beschikken over adequate databestanden waarin per patiënt de betreffende diagnostische en therapeutische gegevens op uniforme wijze zijn opgeslagen. In veel

CAM-praktijken is dat echter nog niet het geval. Een van de grootste struikelblokken bij menig CAM-onderzoek is dat er uiteindelijk veel minder proefpersonen beschikbaar komen dan waarop gerekend was, waarna dan een al lopend onderzoek misschien moet worden stopgezet met alle consequenties van dien. Men komt er dan ook niet omheen voor de start van het onderzoek de potentiële participatie van patiënten bij elk van de deelnemende praktijken te peilen. Dit punt is vooral van belang in situaties waar met controlegroepen wordt gewerkt, maar ook 'outcomes research' is gebaat bij een voldoende grote instroom van patiënten binnen een bepaald tijdsbestek. Niet voor niets wordt de rekrutering van patiënten door CAM-onderzoekers vaak als probleem nummer één ervaren, vooral waar men afhankelijk is van verwijzing door reguliere artsen. De vraag hoeveel proefpersonen je minimaal nodig hebt voor een 'Clinical Outcomes Study' is moeilijk in zijn algemeenheid te beantwoorden. Het is duidelijk dat je met vijftig deelnemers meer te weten komt dan met vijf deelnemers. In het concrete geval kan men het beste uitgaan van de vraag welk aantal personen minimaal haalbaar lijkt binnen een bepaalde context. En die context wordt bijvoorbeeld bepaald door het gemiddelde patiëntenaanbod in de praktijken waar het onderzoek plaats vindt en door de periode die het totale onderzoek mag gaan duren, inclusief voorbereiding, financiering, data-verwerking en rapportage. Als dan is vastgesteld gedurende hoeveel weken of maanden proefpersonen in het onderzoek uiteindelijk kunnen instromen, kan men een voorzichtige schatting maken van het aantal proefpersonen dat uiteindelijk zal meedoen en blijven meedoen.

Het kan natuurlijk zijn dat er van te voren toch behoefte is aan een concreet antwoord op de vraag hoeveel proefpersonen er voor het onderzoek nodig zijn, zeker in het geval dat het verwachte effect van de interventie mogelijk niet zeer groot is. In dat geval is een denkbare richtlijn dat men aan het eind van de behandeling tweehonderd analyseerbare proefpersonen wil overhouden, dat wil zeggen per-

sonen van wie een eindmeting bekend is en die volgens protocol zijn behandeld. Als er dan tijdens het onderzoek om allerlei redenen nog bijvoorbeeld vijftig uitvallen, dan heeft de resterende populatie van honderdvijftig personen, statistisch gezien, nog steeds een aanvaardbare omvang.

Het spreekt vanzelf dat ook wordt aangegeven hoe naar de patiënten toe werd voldaan aan de wettelijke eis van informed consent. Welke informatie ontvingen de (potentiële) proefpersonen? Hoe werd hun anonimiteit gewaarborgd? Konden de patiënten aangeven of ze wel of niet wilden participeren? Is er een vertrouwenspersoon (arts?) beschikbaar waar patiënten tijdens het onderzoek indien gewenst contact mee kunnen opnemen? Soms willen patiënten niet dat hun huisarts op de hoogte is van het feit dat ze aan een (alternatieve) studie deelnemen. Onderzoekers dienen hiermee rekening te houden. Een ander punt is of het onderzoek getoetst werd door een zogenaamde Medisch Ethische Toetsingscommissie (METC). Wettelijk moet tegenwoordig elke wetenschappelijke studie met mensen ter goedkeuring worden voorgelegd aan een METC. In het geval van een niet-gecontroleerde studie lijkt ons die goedkeuring echter niet nodig. Sommigen geven het advies het onderzoeksprotocol van tevoren ter informatie naar een METC op te sturen (White & Ernst, 2001). Zeer betreurenswaardig is het in dit verband dat een speciaal voor het Nederlandse complementaire veld ingestelde METC onlangs moest worden opgeheven omdat er onvoldoende protocollen werden ingediend.

Design (opzet)

Over de keuze voor de opzet, het design, van de studie kunnen we kort zijn. Zoals hierboven al aangegeven, beperken we ons hier tot een onderzoek zonder controlegroepen, een 'Clinical Outcomes Study' (COS). Wel is het goed als de auteur van het artikel enkele woorden wijdt aan de vraag waarom niet voor een gecontroleerde studie werd gekozen. Men kan een paar bekende voor- en nadelen van COS en RCT noemen, zodat de lezer ziet dat de

onderzoeker zich in de materie verdiept heeft en bewust voor 'outcomes research' koos. Heeft de betreffende behandeling tot nu toe bij veel patiënten duidelijk succes dan is dat, zoals hierboven al betoogd, een belangrijk argument voor het doen van een studie zonder (gerandomiseerde en geblindeerde) controlegroepen.

Men kan natuurlijk ook controleren zonder randomisatieprocedure door gebruik te maken van een zogenaamde quasi-experimentele vergelijking, waarbij gewerkt wordt met natuurlijke (niet at random ingedeelde) groepen proefpersonen. Zo kan men bijvoorbeeld in twee simultane studies de genezingspercentages vergelijken van astmapatiënten in een homeopathische en in een conventionele praktijk. We gaan hier niet verder in op de voor- en nadelen van een dergelijke opzet. Ook de toepassing van historische controles en wachtlijstcontroles moeten hier buiten beschouwing blijven.

Heeft men geen behoefte aan een vergelijking met een controlegroep of met een andere setting dan zijn er toch nog andere controlemogelijkheden denkbaar. Zo kan de tijdsduur vanaf het begin van de klachten tot het moment van de interventie worden vergeleken met de tijd die na de behandeling nodig was om te klachten te doen verdwijnen of verminderen. Naarmate het verhoudingscijfer voor/na groter is kan dat een extra aanwijzing zijn voor de werkzaamheid van de behandeling. Deze zogenaamde 'Vorher/nachher Zeit Ratio' (Kiene, 2001) is overigens een niet altijd even betrouwbaar criterium, zo concluderen Alexander & White (2000) in een acupunctuurstudie. Nog een andere controlemaatregel is de zogenaamde 'Patiënt Remedy Score' in de klassieke homeopathie. Na behandeling met een bepaald homeopaticum wordt bekeken of het resultaat van die behandeling correleert met de mate van (door de behandelaar van tevoren vastgestelde) gelijkendheid van het middel. Men toetst met deze procedure de geldigheid van de gelijksoortigheidsregel in de homeopathie: hoe beter het geneesmiddelbeeld van het homeopaticum lijkt op het klachtenbeeld van de patiënt, hoe werk-

zamer het middel (Wiegant et al, 2002). Tenslotte is het ook nog mogelijk – we verlaten dan even de ‘outcomes research’-methode – de patiënt als zijn eigen controle te gebruiken in een zogenaamd ‘single subject design’ ook wel ‘N of 1 design’. Hier confronteert men individuele proefpersonen met een reeks interventies, waarbij fasen van wel en niet behandelen elkaar opvolgen en de effecten daarvan bij die proefpersoon worden geregistreerd. Ook kunnen in een dergelijke opzet fasen met andere therapieën en zelfs placebo’s worden ingelast. Volgens sommigen is het ‘single subject design’ ondanks beperkingen zoals het ‘carry-over effect’, een krachtiger methode dan ‘outcomes research’. Zie verder voor dit onderwerp de specifieke literatuur, bijv. Kiene (2001).

Omdat in ‘outcomes research’ geen controlegroepen worden gebruikt, is blinding voor wie in welke groep terecht komt vanzelfsprekend niet aan de orde. Wel verdient het aanbeveling degene die de data verwerkt te blinderen, in die zin dat hij geen contact heeft met de proefpersonen. Dit kan betekenen dat de onderzoeker het datamanagement aan een derde overlaat, bijvoorbeeld aan een statisticus. Daarbij moet wel worden opgemerkt dat een statisticus die bereid is mee te denken in het procesmatige karakter van ‘outcomes research’ niet gemakkelijk te vinden is.

Betreft het onderzoeksverslag een exploratieve (pilot) studie, dan moet de reden daarvoor duidelijk worden vermeld. Een belangrijk argument voor een pilot is de wens ervaring op te doen met de techniek van het doen van onderzoek.

Experimentele interventie (getoetste behandeling)

Na de beschrijving van het type design dat werd toegepast wordt in detail ingegaan op de diverse fasen in het behandelproces tijdens het onderzoek (interventie). Men geeft het verloop aan van anamnese, lichamelijk onderzoek (ook eventuele foto’s en laboratoriumtesten), diagnose, behandeling en de eventuele follow-up. Verder vermeldt men het

aantal behandelsessies per patiënt en de momenten waarop de evaluatiemetingen worden verricht. Een probleem bij het verslag van het behandelproces is wel dat in menige CAM-praktijk een veelheid aan therapieën toepassing vindt. Zo doen sommige homeopaten ook af en toe aan fytotherapie of aan acupunctuur en de meeste NEI-practitioners praktiseren niet alleen NEI (Neuro-Emotionele Integratie) maar gebruiken ook NLP-technieken of schrijven geneesmiddelen voor, waaronder Bach-remedies en zogenaamde Integra-middelen. Dit betekent dat, willen werkzaamheids-studies enige zeggingskracht hebben, nauwkeurig moet worden vastgelegd wat de behandelaar tijdens het onderzoek bij zijn patiënten aan therapieën of gedragsbeïnvloeding heeft gedaan. Als er meerdere behandelaars aan het onderzoek meedoen, zal hierover van te voren uitvoerig onderling overleg moeten plaats vinden, opdat de therapeutische interventies niet te veel van elkaar afwijken. Waar mogelijk maakt men afspraken over consultfrequenties, doseringen, toedieningsvoorschriften en -tijdstippen, totale duur, etc.. Aan de andere kant moet men er ook weer voor waken te veel in de professionele competentie van de individuele behandelaar in te grijpen en in contraproductieve protocollen te verzanden. Indien bij bepaalde proefpersonen tijdens het onderzoek (gedeeltelijk) van de geplande interventie werd afgeweken, dient dat ook vermeld te worden.

Een tekening van het geplande tijdpad voor het gehele interventieproces kan heel verhelderend zijn voor de lezer. Figuur 1 geeft een eenvoudig voorbeeld van een tijdpad dat elke proefpersoon geacht wordt te doorlopen. Het werd ontleend aan het voorlopige onderzoeksverslag van het al eerder genoemde osteopathieonderzoek. Uit de figuur blijkt dat er drie behandelsessies waren met een tussenruimte van een week. Verder ziet men wanneer de twee vragenlijsten moesten worden ingevuld. De eerste lijst moest twee keer worden ingevuld, aan het begin en aan het eind van de studie; de tweede lijst alleen aan het eind. In onderzoeksvoorstellen en definitieve protocollen wordt meestal een veel

gedetailleerder getekend tijdpad opgenomen, inclusief de fasen van voorbereiding en verslaglegging. Tenslotte is het soms handig het complete studiedesign van te voren te publiceren (in een tijdschrift of op internet) met het verzoek aan de lezers om commentaar.

Uitkomstmaten (variabelen)

De in het onderzoek gebruikte uitkomstmaten of variabelen werden, als het goed is, al in de inleiding van het artikel bij de vraagstelling vermeld. Bij het onderzoek naar het effect van osteopathie bij huilbaby's was de primaire uitkomstmaat het aantal huiluren overdag en 's nachts. Dit is een voorbeeld van een zogenaamde objectieve maat, uit te drukken in een getal. Andere objectieve variabelen zijn bloeddruk, bloed- en urinewaarden, visus, huidploidikte, 1-sec. waarde, etc.. Hiertegenover staan subjectieve uitkomstmaten, zoals pijnbeleving, tevredenheid en algemeen welbevinden. Bij het onderdeel 'resultaten' wordt nog wat meer gezegd over uitkomstmaten.

Dataverwerking

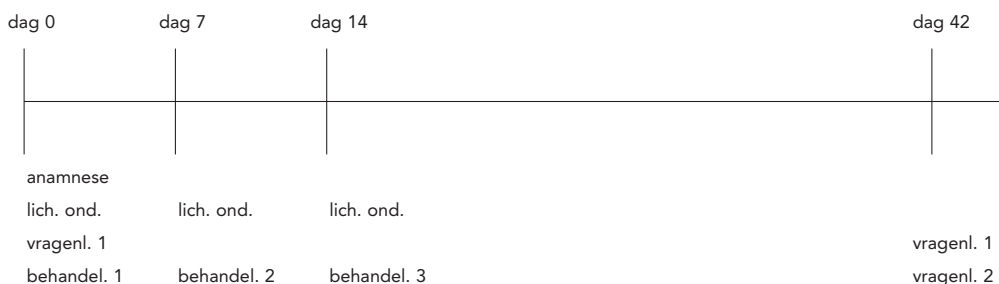
Na een beschrijving van de uitkomstmaten of variabelen zal het artikel logischerwijze ingaan op de vraag: hoe is de onderzoeker aan zijn data gekomen? De auteur zal hierin duidelijk aangeven welke

meetinstrumenten hij gebruikt heeft en hoeveel metingen hij daarmee per proefpersoon deed. Terzijde zij hier opgemerkt dat het beslist niet altijd nodig is voor de beantwoording van een onderzoeksvraag nieuwe data te verwerven. We noemden al de mogelijkheid proefpersonen retrospectief te rekruteren, waarbij men dan gebruik maakt van in het verleden geregistreeerde patiëntgegevens. Maar men kan ook bevolkingscijfers raadplegen die door officiële instanties worden bijgehouden, bijvoorbeeld ziekte- en sterftetafels.

Wat de keuze voor het meetinstrument betreft, het moge duidelijk zijn dat die vaak direct voortvloeit uit de vraagstelling waarin de uitkomstmaat werd vastgelegd. Zo zal een uitkomstmaat 'bloeddruk' gemeten worden met behulp van het instrument 'bloeddrukmeter'. En de longfunctie meet men met een 'peak flow meter'. Soms ook komen er nieuwe meetinstrumenten in de handel, die dan in onderzoek nog geijkt en gevalideerd moeten worden. Op het gebied van CAM-behandelingen wordt veel gebruik gemaakt van de vragenlijst als meetinstrument. Vaak is dat daar ook de enige mogelijkheid. Het invullen van vragenlijsten kan op allerlei manieren gebeuren, bijvoorbeeld door de proefpersoon thuis. Men krijgt de vragen toegestuurd dan wel op de praktijk uitgereikt. Invullen ter plaatse in het bijzijn van of zelfs door de onderzoeker resp. behan-

Figuur 1

Onderzoek osteopathische behandeling bij huilbaby's: interventie-tijdpad per baby



delaar kan ook. De wijze waarop de lijsten worden verstuurd, uitgereikt, ontvangen en verwerkt moet goed van te voren worden doordacht en gepland. Dat geldt ook voor de manier waarop men niet-reagerende respondenten eventueel opnieuw benadert.

De constructie van een vragenlijst is geen sinecure en eigenlijk een vak apart. Een goede lijst moet aan allerlei eisen voldoen. Zo moeten de verschillende onderwerpen (items) zorgvuldig en ondubbelzinnig worden geformuleerd en in een geschikte volgorde op papier komen. Ook moet een vragenlijst een heldere lay-out hebben en het lezen en invullen mag niet te veel tijd vragen. Proefpersonen moeten gemotiveerd worden om serieus met de lijst om te gaan. Het is aan te bevelen een vragenlijst eerst bij collega's en kennissen uit te testen. Een belangrijk discussiepunt onder deskundigen, waar we hier verder niet op ingaan, is de zogenaamde validiteit van vragenlijsten: meet je eigenlijk wel wat je denkt te meten. Er is veel literatuur over het maken van vragenlijsten voor wetenschappelijk onderzoek. Een beknopte inleiding is o.a. te vinden bij Brinkman (1998).

Dataverwerking (statistische analyse)

Hier komen we op het punt waar we het niveau van een al of niet systematische praktijkregistratie gaan overstijgen. We gaan nu wat doen met de verworven gegevens. In principe kan een registratie van de lotgevallen van patiënten eindeloos worden voortgezet, maar op een gegeven moment zal men toch een idee willen hebben van wat de verzamelde gegevens betekenen of anders gezegd: wat de behandeling eigenlijk oplevert. Dat betekent dat de gegevens nader moeten worden geanalyseerd.

De belangrijkste vraag die we ons hierbij stellen luidt: kan een afname van de klachten worden toegeschreven aan de onderzochte behandeling (interventie) of is er toeval in het spel. Voor een antwoord op deze vraag kan gebruik gemaakt worden van zogenaamde significantietoetsen. Het is hier niet de plaats op de techniek van dergelijke toetsen

in te gaan en het is dan ook voor de doorsnee-onderzoeker beslist noodzakelijk methodologische en statistische deskundigheid in te huren als hij die niet zelf in huis heeft. Vooral een bio-statisticus kan de onderzoeker voor veel valkuilen behoeden. Een van die valkuilen betreft de potentiële afhankelijkheid van herhaalde waarnemingen, waar we hier verder niet op in kunnen gaan.

Tegenwoordig wordt de computer ingeschakeld bij het doen van toetsen. Het bekendste statistische programma met veel mogelijkheden is hier SPSS. In het onderzoeksverslag is het niet nodig de achtergronden van de gebruikte statistische toetsen te behandelen. Men mag ervan uitgaan dat de lezer daar voldoende van weet. Wel is het zaak duidelijk aan te geven welke toetsen werden toegepast en waarom.

Enkele middelen om onderzoeksresultaten samenvattend te beschrijven

a frequentieverdelingen

- tabel
- diagram
- grafiek

b centrummaten

- gemiddelde
- mediaan

c spreidingsmaten

- variantie
- standaarddeviatie

d uitkomsten van statistische toetsen

Vierde onderdeel van het artikel: resultaten

Dit onderdeel geeft antwoord op de vraag: 'what did you find?' Welke uitkomsten heeft het onderzoek opgeleverd? De verschillende onderzoeksresultaten worden gepresenteerd in een samengevatte vorm, dat wil zeggen met relevant cijfermateriaal en toelichtende teksten. De relevante cijfers worden gepresenteerd in zogenaamde frequentieverdelingen annex centrummaten, spreidingsmaten en resultaten van significantietoetsen.

Frequentieverdelingen zijn tabellen en figuren (diagrammen, grafieken) waarin men in één plaatje de verschillende meetwaarden van een variabele kan aflezen, bijvoorbeeld de leeftijdsverdeling van een groep proefpersonen. Centrum- en spreidingsmaten geven in één getal een beeld van de gemeten variabele, bijvoorbeeld de gemiddelde leeftijd (een centrummaat) van de proefpersonen of de standaarddeviatie van de leeftijden (een spreidingsmaat). De standaarddeviatie geeft een idee van de mate waarin de individuele leeftijden afwijken van de gemiddelde leeftijd.

Overigens moeten, ideaal gezien, dit soort datapresentaties voor zichzelf spreken. Men zou zich dan ook als auteur bij elk plaatje en bij elke cijfer moeten afvragen: snap ik deze tabel of grafiek als zodanig, ook zonder de toelichting in de tekst? Bij de behandeling van het thema 'response data' hieronder geven we een voorbeeld van de manier waarop gemiddelden, standaarddeviaties en statistische significantie in een tabel kunnen worden opgenomen; zie tabel 3 verderop in deze bijdrage.

We spraken hierboven niet voor niets over *relevante* cijfers. Minder belangrijke data die het onderzoek heeft opgeleverd, worden in een tijdschriftartikel namelijk zoveel mogelijk weggelaten, ook al omdat tijdschriftredacties meestal zeer beperkt (cijfermatige) illustraties opnemen. Ze kunnen wel als bijlagen worden meegenomen in een uitgebreid onderzoeksrapport.

Welke soort gegevens in de rapportage terechtkomt hangt primair af van de vraagstelling en de soort

studie. In het algemeen registreert men in een werkzaamheidsstudie voor elke proefpersoon drie soorten patiëntgegevens (zie kader): prognostische patiëntgegevens (a), behandel- of interventiegegevens (b) en gegevens over de reactie op de behandeling of 'response data' (c). De prognostische gegevens zijn globaal te verdelen in demografische (o.a. geslacht, leeftijd) en medische gegevens van de patiënt (voorgeschiedenis, huidige klachten, etc.).

Patiëntgegevens die tijdens een klinische studie worden geregistreerd

a prognostische patiëntgegevens

- demografische gegevens
- medische gegevens

b behandel- of interventiegegevens

c 'response data'

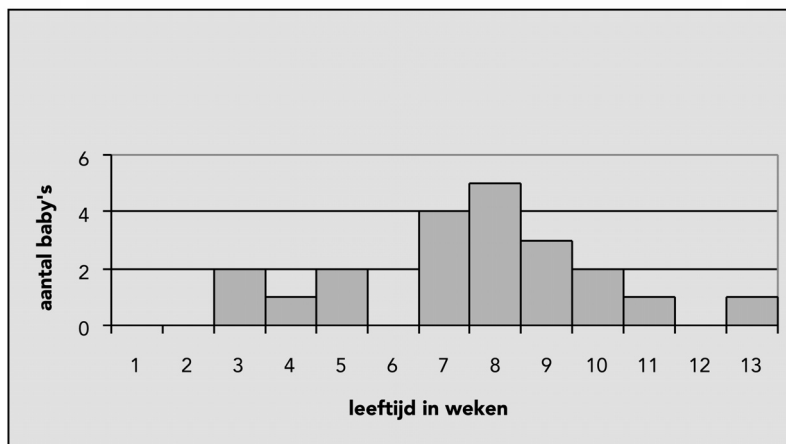
Prognostische patiëntgegevens

Tabel 1 is een voorbeeld van de manier waarop de verschillende prognostische patiëntgegevens op eenvoudige wijze kunnen worden weergegeven. Figuur 2 is een voorbeeld van een veel gebruikte frequentieverdeling, het histogram. In één oogopslag ziet men hoe de leeftijden van proefpersonen in een bepaalde studiepopulatie zijn verdeeld. Beide (aangepaste) voorbeelden werden ontleend aan de al eerder genoemde osteopathiestudie. Men moet wel bedenken dat het slechts voorbeelden zijn en geen standaardindelingen. Zo is in tabel 1 het samenvoegen van de aantallen huiluren in twee categorieën (overdag: 3-8 en >8; 's nachts: 0-2 en >2) arbitrair. Is het aantal proefpersonen groter dan in deze voorbeelden, dan is het nuttig om (ook) percentages te weer te geven.

Tabel 1

'Clinical Outcome Study' naar het effect van een osteopathische behandeling bij huilbaby's; frequentie van enkele prognostische patiëntgegevens

prognostische patiëntgegevens huilbaby's		aantal (n=21)
<u>geslacht:</u>	man	7
	vrouw	14
<u>leeftijd:</u>	3-7 wk	5
	7-14 wk	16
<u>medische voorgeschiedenis:</u>	ziekenhuisopname wegens huilen	2
	etc.	?
<u>eerdere interventies:</u>	medicijngebruik tijdens zwangerschap	5
	kunstgrepen bij bevalling	12
	borstvoeding	15
	inbakeren	20
	etc.	?
<u>huidige klachten en symptomen:</u>	3-8 huiluren overdag	19
	> 8 huiluren overdag	2
	0-2 huiluren 's nachts	13
	> 2 huiluren 's nachts	8
	etc.	?



Figuur 2
Leeftijd baby's bij
begin behandeling
(N=21)

Behandelgegevens

Registratie van de tweede soort data, de behandel- of interventiegegevens per patiënt, is vooral van belang voor gecontroleerde studies waar men zo nauwkeurig mogelijk bijhoudt welke proefpersonen uiteindelijk wel en niet de te toetsen behandeling (experimentele interventie) ondergingen. Om de verschillende patiëntengroepen (therapie- en controlegroepen) te kunnen vergelijken zal men immers na afloop willen weten hoeveel mensen in welke groep om welke reden dan ook, niet de voorgeschreven standaardbehandeling kregen (problemen van 'ineligible patients', 'non-compliance', 'withdrawals', 'drop-outs'). In het geval van een interventieperiode met meerdere behandelingsessies houdt de onderzoeker per proefpersoon bij of er afwijkingen optraden in het geplande sessieritme. Als iemand bijvoorbeeld twee keer niet komt opdagen en daarna weer wel, geeft men aan of de behandeling gewoon werd voortgezet of dat er opnieuw is gestart met de officiële interventie.

Voor een adequate vergelijking hanteert men meestal het 'intention to treat' principe, dat wil zeggen dat alle proefpersonen inclusief de uitvallers worden meegenomen in de statistische analyse. Maar ook in een 'Clinical Outcomes Study' (COS) is het 'intention to treat' principe van belang en zal men zo nauwkeurig mogelijk registreren hoe de behandeling bij de deelnemers verloopt. Want ook

hier wil men een duidelijk beeld verkrijgen van de lotgevallen van de studiepopulatie. Niet in de laatste plaats vanwege het verschijnsel van de geleidelijke 'uitdunning' van de studiepopulatie: de ervaring is dat patiënten soms al na een of twee consulten wegblijven, bijvoorbeeld om financiële redenen of omdat ze geen tijd of zin meer hebben. In de daarop volgende sessies is de belangrijkste reden voor uitval dat de patiënt zich zodanig verbeterd voelt dat hij afziet van een verdere behandeling. Verder is het zo dat afhankelijk van de soort therapie het gros van de patiënten toe kan met een relatief gering aantal behandelingen, bijvoorbeeld vier tot zes, terwijl therapie-resistente patiënten bijvoorbeeld tien of meer sessies nodig kunnen hebben.

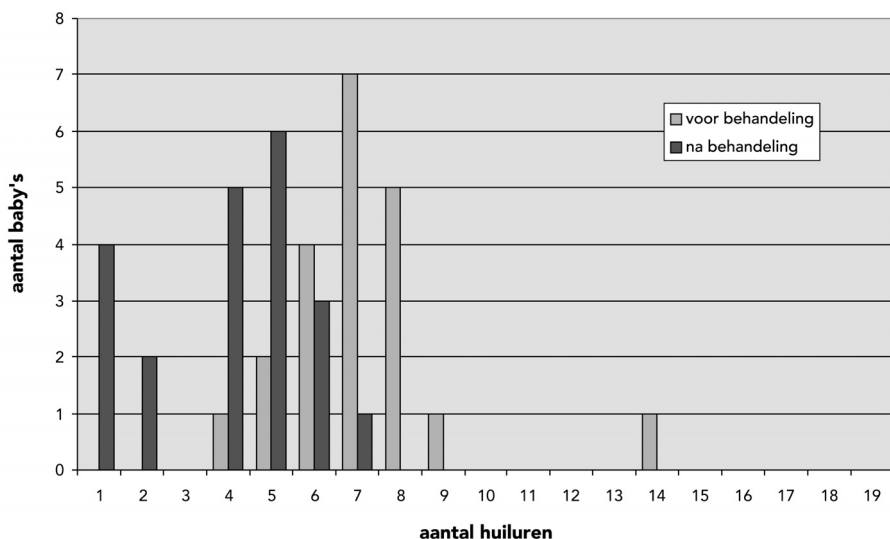
'Response data'

En dat brengt ons op de derde categorie data: de 'response data'. Dat zijn gegevens over de reactie (response) van de proefpersonen op de getoetste behandeling. In het eenvoudigste geval hebben we dan te maken met twee categorieën: wel en niet vermindering van de klachten of symptomen. Een meer gedifferentieerde, oplopende classificatie is bijvoorbeeld: de pijn is totaal weg, is gedeeltelijk weg, is niet veranderd, is verergerd. Om dergelijke gegevens te verkrijgen zou men proefpersonen, die bijvoorbeeld een acupunctuurbehandeling ondergingen, na afloop een enquêtevraag kunnen voor-

Tabel 2

Door migrainepatiënten gepercipieerde hoofdpijnklachten voor en na een acupunctuurbehandeling

	voor	na
weinig tot geen hoofdpijn	0 (0%)	30 (18%)
milde hoofdpijn die ik af en toe kan verdringen	48 (28%)	69 (41%)
behoorlijke hoofdpijn maar ik kan mijn normale werk blijven doen	66 (39%)	33 (20%)
intense hoofdpijn die mij compleet in de greep heeft	54 (32%)	36 (21%)
totaal aantal patiënten	168 (100%)	168 (100%)



Figuur 3
aantal huiluren
overdag voor en na
een osteopathische
behandeling bij
huilbaby's (N=21)

leggen met de hierboven genoemde vier antwoordmogelijkheden. De patiënt wordt dan gevraagd in gedachten terug te gaan naar de pijnbeleving voor de behandeling om deze dan te vergelijken met de pijnbeleving na de behandeling. Het is echter in onderzoek gebruikelijk de zaak wat 'objectiever' aan te pakken en niet alleen een meting maar ook een voormeting te doen, ook al omdat patiënten vaak niet meer precies weten hoe erg hun klachten voor de interventie waren. Zie tabel 2 voor een hypothetische studie waarin migrainepatiënten op een 4-puntsschaal konden aangeven hoe erg de hoofdpijn was voor en na een acupunctuurbehandeling. Behalve in een tabel kunnen de cijfers ook op andere wijze worden gepresenteerd, bijvoorbeeld in een cirkel- of taartdiagram.

Voor- en nametingen kunnen heel simpel of heel geavanceerd van karakter zijn. Men kan zich beperken tot één meetinstrument, maar men kan ook gebruik maken van verschillende instrumenten en van meerdere metingen per proefpersoon. Zo kan de patiënt gevraagd worden bovenstaande 4-puntsschaal niet twee keer maar veel vaker in te vullen, bij wijze van dagboek. Zo kan men dan bijvoorbeeld vier maal per dag de mate van hoofdpijn registre-

ren. Soms wordt een dergelijk dagboek een aantal perioden bijgehouden, bijvoorbeeld gedurende drie perioden van vier weken: vóór het onderzoek als baseline, drie maanden erna en een jaar later als follow-up.

Behalve in specifieke klachten kan de onderzoeker ook geïnteresseerd zijn in veranderingen in het algemeen welbevinden van de proefpersonen. Voor dit aspect zijn verschillende (gevalideerde) vragenlijsten in omloop.

Naast de hierboven besproken *kwalitatieve* gegevens levert wetenschappelijk onderzoek ook veel *kwantitatieve* gegevens waar allerlei berekeningen op kunnen worden losgelaten, zoals het bepalen van gemiddelde waarden. Als we ouders van huilbaby's in een enquête vragen naar het aantal uren dat de baby huilt voor en na een behandeling, krijgen we een heel ander soort data dan in het geval van de pijnstudie hierboven. We vragen de proefpersonen nu niet naar 'subjectieve' oordelen; we vragen ze naar 'objectieve' cijfers, een cijfer van voor en een cijfer van na de behandeling. De aantallen huiluren die uit de enquête komen kunnen op verschillende manieren in de rapportage worden gepresenteerd. We zouden om te beginnen een fre-

quentietabel kunnen maken waarin je kunt zien hoeveel baby's hoeveel uren huilen, en hoeveel uren er in totaal voor en na de behandeling gehuild werd. Wanneer er veel proefpersonen zijn, is het vaak overzichtelijker een figuur te presenteren in plaats van een tabel, bijvoorbeeld een histogram. In figuur 3 zien we een histogram met de frequentieverdeling van de aantallen huiluren voor en na de behandeling.

Het is ook mogelijk om je in een artikel te beperken tot een tabel met uitsluitend een centrummaat annex spreidingsmaat en het resultaat van een significantiebepaling. Tabel 3 is hier een voorbeeld van. De gegevens werden ontleend aan een placebocontroleerde acupunctuurstudie (Nabeta et al, 2002). Omdat de resultaten van de placebobehandeling voor ons verhaal niet relevant zijn lieten we die weg. Voor drie verschillende locaties werd bij een verschillend aantal proefpersonen met een bepaald instrument de pijn drempel (PPT) gemeten voor en na de behandeling. Als centrummaat werd het gemiddelde genomen en als spreidingsmaat de standaarddeviatie. Tenslotte werd met een zogenaamde *t*-toets berekend of de verschillen tussen de gemiddelde pijn drempelwaarden (voor en na) significant waren. Uit de in de tabel vermelde *p*-waarde ($p < 0.001$) blijkt dat die significantie inderdaad bestond, dat wil zeggen dat de geconstateerde verschillen hoogstwaarschijnlijk niet op toeval

berustten.

Wat we noch uit figuur 3 noch uit tabel 3 kunnen opmaken is het *aantal* proefpersonen dat baat heeft gehad bij de therapie, we zien alleen gemiddelde cijfers. De cijfers in tabel 3 laten ons niet zien bij hoeveel patiënten de pijn drempel daalde dan wel steeg, we zien alleen dat hij gemiddeld daalde. En het histogram in figuur 3 geeft ons geen antwoord op de vraag hoeveel baby's in totaal minder (dan wel meer) zijn gaan huilen na de behandeling. En het is zeker belangrijk ook deze gegevens te laten zien (in tabellen). Want het kan best zijn dat het totaal aantal huiluren na de osteopathiebehandeling aanzienlijk gedaald is, maar dat het percentage baby's dat minder is gaan huilen toch tegenvalt. Het statistische computerprogramma SPSS maakt het ons gelukkig tegenwoordig mogelijk, mits alle data adequaat werden ingevoerd, al dit soort cijfers in een handomdraai in tabellen en figuren te produceren.

Als het onderzoek beperkt van omvang is en minder dan 20 à 30 proefpersonen omvat, dan kan men de exacte reactie per patiënt ook *grafisch* weergeven; zie figuur 4, waar de verandering van het aantal huiluren bij een selectie van drie baby's direct te zien is. Het is dan ook meteen duidelijk welke en hoeveel proefpersonen er wel en niet op vooruit gingen. In onze eerder geformuleerde voorbeeld-vraagstelling werd overigens niet gevraagd naar de lotge-

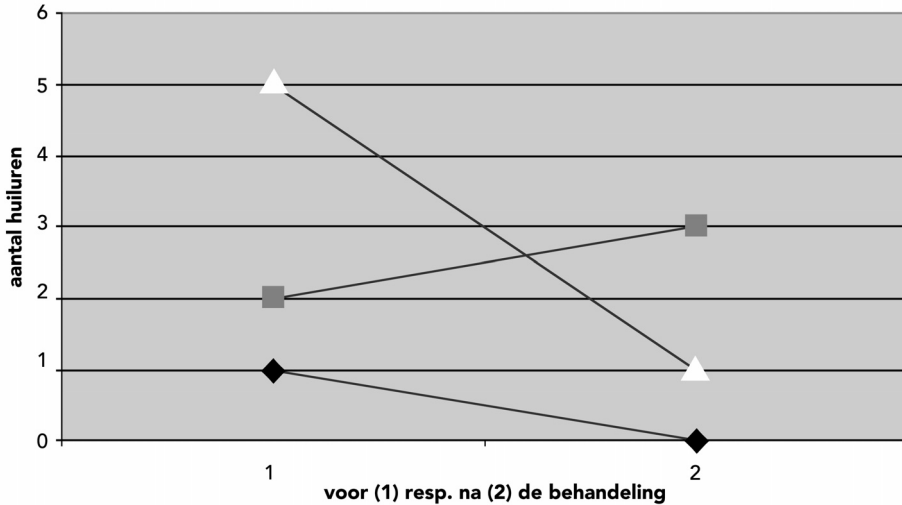
Tabel 3

Pressure pain threshold (PPT) gemeten voor en na een acupunctuurbehandeling

	n	voor	na
nek	27	1.7 ± 0.7*	2.1 ± 0.9**
schouder	45	2.4 ± 1.8*	2.9 ± 2.3**
rug	24	2.3 ± 0.8*	2.7 ± 1.1**

* gemiddelde ± standaarddeviatie

** $p < 0.001$, gepaarde *t*-toets voor vs. na

**Figuur 4**

Aantal huiluren per baby (N=3) voor en na de behandeling

vallen per baby. Men zou dan ook kunnen overwegen de vraagstelling op dit punt te concentreren of beide elementen erin op te nemen, zowel de verandering van het aantal huiluren over de totale populatie als het huilgedrag per baby.

Vijfde onderdeel van het artikel: discussie

In het onderdeel 'discussie' tenslotte geeft de onderzoeker zijn eindconclusie, gebaseerd op de resultaten. Hij geeft antwoord op de vraag: 'what does it mean?' Bij zijn conclusie betreft hij nadrukkelijk de vraagstelling zoals die in de inleiding werd geformuleerd. Verder is een algemeen uitgangspunt dat de onderzoeker-auteur om zijn wetenschappelijke geloofwaardigheid te behouden de nodige terughoudendheid betracht bij het trekken van conclusies.

In het geval van onze voorbeeld-vraagstelling zou de conclusie kunnen luiden – bij statistische significantie van de resultaten – dat de osteopathische behandeling zoals toegepast door de genoemde behandelaar in de genoemde setting een gunstig effect lijkt te hebben op overmatig huilgedrag bij

baby's. Een dergelijke conclusie wint aan kracht als een eventuele positieve uitkomst van het onderzoek tevens klinisch relevant is, hetgeen in dit geval zou inhouden: het aantal huiluren is bij minstens de helft van de baby's met 50% afgenomen.

Na het trekken van een (voorlopige) conclusie kan de auteur verder de sterke en zwakke kanten van het onderzoek belichten. Tenslotte kunnen er aanbevelingen voor de praktijk volgen en suggesties voor vervolgonderzoek.

Literatuur

- Alexander, R., White, A. (2000) Acupuncture in a rheumatology clinic, *Acupunct Med*, 18, pp.100-103
- Batstra, J.T. (2004) *Het effect van een osteopathische behandeling bij huilbaby's*, Thesis, Utrecht [geen officiële publicatie]
- Brinkman, J. (1998) *Cijfers spreken. Statistiek en methodologie voor het hoger onderwijs*, Groningen, Wolters-Noordhoff
- Kiene, H. (2001) *Komplementäre Methodenlehre der klinischen Forschung, Cognition-based Medicine*, Berlin, Springer

- Nabeta, T., Kawakita, K. (2002) Relief of chronic neck and shoulder pain by manual acupuncture to tender points – a sham-controlled randomized trial, *Comp Ther Med*, 10, pp.217-222
- Paton, A. (1979) geciteerd door Pocock SJ. (1987, 1983), *Clinical trials, a practical approach*, Chichester, Wiley
- Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., Richardson, W.S. (1996) Evidence based medicine: what it is and what it isn't, *Br Med J*, 312, pp.71-72
- Vickers, A.J. (2002) Inspiration and perspiration: what every researcher needs to know before they start, In: Lewith, G., Jonas, W.B., Walach, H. (Eds) *Clinical Research in Complementary Therapies, principles, problems and solutions*, Edinburgh Churchill Livingstone, pp.47-58
- Walach, H., Jonas, W.B., Lewith, G. (2002) The role of outcomes research in evaluating complementary and alternative medicine, In: Lewith, G., Jonas, W.B., Walach, H. (Eds) *Clinical Research in Complementary Therapies, principles, problems and solutions*, Edinburgh Churchill Livingstone, pp.29-45
- White, A., Ernst, E.. (2001) The case for uncontrolled clinical trials: a starting point for the evidence base for CAM, *Comp Ther Med*, 9, pp.111-115
- Wiegant, F., Kramers, C.W., Wijk, R. van. (2002) The importance of patient selection, In: Lewith, G., Jonas, W.B., Walach, H. (Eds) *Clinical Research in Complementary Therapies, principles, problems and solutions*, Edinburgh Churchill Livingstone, pp.155-169

> SamenvattingClinical Outcomes Studies voor CAMOpzet en verslaglegging

Onderzoek zonder controlegroepen of 'outcomes research' biedt een goede mogelijkheid voor het toetsen van de werkzaamheid van CAM-behandelingen. De gebruikelijke gecontroleerde studies (RCT's) komen, onder meer vanwege een aantal ethische en praktische bezwaren, op dit moment minder in aanmerking. In deze bijdrage wordt ingegaan op de

methodologie van een ongecontroleerde 'Clinical Outcomes Study'. Opzet en rapportage worden toegelicht vanuit de positie van de onderzoeker-auteur die aan het eind van het onderzoek verslag moet doen van de resultaten. Stapsgewijs wordt het opzetten van het onderzoek en de verslaglegging doorgenomen.

> SummaryClinical Outcomes Studies for CAMAspects of planning and reporting

'Outcomes research' can have a pivotal role in CAM-research; this approach does not require the use of control groups. In contrast to randomized clinical trials 'outcomes research' has little methodological,

ethical and practical limitations. Aspects of planning and reporting will be reviewed. Organising the research and publishing the results is presented step by step.

Key words

CAM ■ clinical trial methodology
■ outcomes research ■ reporting of trial results

Auteur

Dr Willem Kramers is arts-epidemioloog en was verbonden aan de universiteiten van Utrecht, Heidelberg (D) en Witten/Herdecke (D). Thans vervult hij de functie van professor voor Health Sciences aan het 'Dutch University College' te Deventer, met als aandachtsveld 'CAM-therapies research and education'.

ADRES

Herenlaan 38
3701 AV Zeist
T +31 (0)30 692 22 67
F +31 (0)30 692 22 67
E c.w.kramers@tele2.nl