

De meta-analyse als review-methode

GERBEN TER RIET
JOS KLEIJNEN
PAUL KNIPSCHILD

Een 'meta-analyse' is een systematische samenvatting van de literatuur over een bepaald onderwerp, inclusief een beschrijving van gebruikte methoden om de literatuur op te sporen en te selecteren. Via Medline en enkele andere strategieën is geprobeerd een zo compleet mogelijk overzicht te krijgen van de beschikbare literatuur over acupunctuur bij een achttal indicatiegebieden: chronische pijn, pijnklachten in en om de wervelkolom, reumatoïde arthritis, migraine en spanningshoofdpijn, aangezichtspijn, astma, stoppen met roken, en alcohol en drugs. De opgespoorde literatuur is vervolgens geselecteerd op basis van twee criteria: het onderzoek moest betrekking hebben op mensen, en er moest een controlegroep bij betrokken zijn. De aldus geselecteerde studies zijn beoordeeld aan de hand van een uitvoerige criterialijst, waarmee maximaal 40 punten konden worden gescoord. Door de uitvoerige toelichting kan de lezer eventueel zelf andere gewichten aan de criteria toekennen en zijn eigen puntenlijst maken.

Ter Riet G, Kleijnen J, Knipschild P. De meta-analyse als review-methode [De effectiviteit van acupunctuur]. Huisarts Wet 1989; 32(5): 176-81

Vakgroep Epidemiologie/Gezondheidszorgonderzoek, Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht.

G. ter Riet, arts; J. Kleijnen, arts; prof. dr. P. Knipschild, hoogleraar epidemiologie.
Correspondentie: G. ter Riet.

Inleiding

Veel literatuurreviews zijn onbevredigend, doordat de schrijvers zeer selectief studies weglaten zonder dat zij melding maken van de criteria die zij daarbij hebben gehanteerd. Een 'meta-analyse' is een systematische samenvatting van de literatuur over een bepaald onderwerp, inclusief een beschrijving van gebruikte methoden om de literatuur op te sporen en te selecteren.

Sommige auteurs reserveren de term 'meta-analyse' voor reviews waarin de studieresultaten statistisch worden gecombineerd ('gepooled'). Hieraan liggen echter vele veronderstellingen ten grondslag, waaraan bijna nooit wordt voldaan, en zeker niet bij het onderwerp acupunctuur. Ter onderscheiding van deze statistische methode wordt daarom wel de term 'op criteria gebaseerde' meta-analyse gebruikt. Het belangrijkste kenmerk van deze methode is de nadruk die wordt gelegd op correcte onderzoeksmethodieken. Een bevredigende methode om resultaten uit verschillende onderzoeken te comprimeren tot één samenvattend getal is ons inziens nog niet voorhanden.

In deze serie over de effectiviteit van acupunctuur is gekozen voor de op criteria gebaseerde meta-analyse. In acht artikelen zullen de volgende indicatiegebieden aan de orde komen: chronische pijn, pijnklachten in en om de wervelkolom, reumatoïde arthritis, migraine en spanningshoofdpijn, aangezichtspijn, astma, stoppen met roken, en alcohol en drugs. In deze tweede inleidende bijdrage wordt uiteengezet hoe deze reviews tot stand zijn gekomen.

Literatuurverzameling

Voor het verzamelen van de literatuur zijn drie zoekstrategieën gecombineerd:

- Op 30 april 1988 is een Medline computersearch gedaan, waarbij de volgende trefwoorden werden gebruikt: *acupuncture, clinical trials, migraine, vascular headache, headache, tobacco, smoking, asthma, rheumatoid arthritis, backache, osteoarthritis, cervical vertebrae, pain and therapy*. Via deze zoekprocedure werden artikelen tot 1963 opgespoord. Via een tweede en een derde search – met gebruikmaking van het

CD-ROM-systeem – werden alle artikelen uit 1988 waarin *acupuncture* was vermeld, opgespoord; deze artikelen werden vervolgens beoordeeld op de insluitcriteria aan de hand van de titel en de samenvatting.

- Via correspondentie met een Franse collega werd de verzameling aangevuld met Franstalige studies. Reizen naar Aken, Marburg en Parijs werden ondernomen ter uitsluiting c.q. opvulling van hiaten.
- Tenslotte werden de opgespoorde artikelen gescreend op bruikbare referenties. In dat kader past ook het screenen van Excerpta Medica (vanaf 1981; indexwoord: *acupuncture*) en van het attenderingsbulletin voor de 'grijze' literatuur (vanaf nummer 3, 1986) van het Nationaal Informatie- en Documentatiecentrum Alternatieve Geneeswijzen te Amersfoort.

Er werden twee insluitcriteria gehanteerd:

- het artikel moest betrekking hebben op onderzoek bij *mensen*; dit werd een klinische studie genoemd;
- bij het onderzoek moest een *controlegroep* zijn betrokken.

Deze criteria sluiten automatisch alle dierexperimentele onderzoeken uit. Alle artikelen waarin series patiënten worden gepresenteerd (voor-na studies) vielen ook af. Veel van dit soort artikelen werden wel gelezen en een groot aantal staat ook vermeld in de literatuurlijsten; zij zijn echter niet gescoord op de criterialijst.

Definities

We zijn uitgegaan van een pragmatische definitie van acupunctuur: er moest sprake zijn van naalden in enige vorm. Dat betekent dat studies waarin gebruik werd gemaakt van verblijfsnaalden of druknaalden ('studs') werden opgenomen. Studies waarin uitsluitend plakelektroden werden gebruikt, werden uitgesloten; dat geldt dus voor studies waarin uitsluitend transcutane elektrische zenuwstimulatie (TENS) werd gebruikt. Studies waarin uitsluitend sprake was van laseracupunctuur werden eveneens uitgesloten. Onze indruk is dat gecontroleerd onderzoek naar de effectiviteit van deze methode erg schaars is. Enkele studies waarin met naalden werd geprikt in een pijnlijke plaats (*locus dolendi*), zonder re-

kening te houden met beschreven acupunctuurpunten, werden wél opgenomen, evenals vormen van elektrostimulatie via naalden (elektro-acupunctuur).

In twee gevallen spreken we van placebo-acupunctuur:

- wanneer met naalden (of druknaalden) wordt gemanipuleerd in of op huidpunten die wel als acupunctuurpunten beschreven staan, maar voor de betreffende ziekte/aandoening niet ter zake doen (placebo-acupunctuur in of op niet-specifieke punten);
- wanneer met naalden (of druknaalden) wordt gemanipuleerd in of op huidpunten die niet als acupunctuurpunten beschreven staan en dus 'gewone' huidpun-

ten zijn (acupunctuur in of op willekeurige huidpunten; *random insertions*).

Het simuleren van elektrostimulatie door niet werkende TENS-apparatuur wordt niet zonder meer als placebo-acupunctuur geaccepteerd.

Het begrip 'placebo-acupunctuurpunt' (kortweg placebopunt) werd analoog aan het begrip placebo-acupunctuur gedefinieerd.

Een acupuncturist die zijn vak traditioneel beoefent, zal bezwaar maken tegen de definitie die hier van placebo-acupunctuur wordt gegeven, omdat deze uitgaat van acupunctuurpunten die specifiek zouden zijn voor bepaalde ziekten of aandoeningen. De traditionele acupuncturist zou

werken met patiënt-specifieke combinaties van punten. Niettemin wordt ook binnen de traditionele stroming gebruik gemaakt van basispatronen van punten bij bepaalde aandoeningen.¹ Ietwat denigrerend wordt dit wel recept- of formule-acupunctuur genoemd. In de praktijk is formule-acupunctuur echter de overheersende stroming, in ieder geval in de westerse wereld. Bijna alle gecontroleerde klinische studies die aan de orde zullen komen, gaan uit van formule-acupunctuur. Dit betekent dat onze conclusies betrekking zullen hebben op de acupunctuur zoals die wordt uitgevoerd door het gros van de beoefenaren op het westelijk halfrond en in Australië en Hong Kong.

De criterialijst

Ter verduidelijking is hier de criterialijst uit het artikel over chronische pijn afgedrukt (tabel). Deze lijst is gebruikt voor alle reviews en is in principe vergelijkbaar met de bekende APGAR-score uit de obstetrie.

Ook op onze criterialijst kunnen punten worden verdiend, maximaal 40. Dit maximum is uiteraard vrij willekeurig en met het toenemen van de kennis over onderzoeksmethodieken zou men in de toekomst wellicht 100 punten te verdelen hebben. Omdat de criteria vooral nadruk leggen op een juiste onderzoeksmethodiek, leidt hun hantering tot uitspraken over de mate waarin men de gevonden resultaten mag vertrouwen.

Vergelijkbaarheid van de groepen

Het begrip *homogeniteit* is zeer problematisch. Het is in de lijst opgenomen om te benadrukken dat men moet streven naar het selecteren van groepen patiënten die met betrekking tot de prognostisch relevante kenmerken op elkaar lijken. Wanneer men patiënten tot een onderzoek toelaat met allerlei klachten en kwalen, is de kans groot dat het opsporen van therapie-effecten bij een subgroep (met specifieke klachten) mislukt. Dit wordt veroorzaakt doordat de resultaten op groepsniveau worden geanalyseerd. Er treedt dan 'verdunding' van eventuele effecten op. Hetzelfde geldt voor de selectie van patiënten die qua ziektestadium erg verschillen. Positieve effecten van acupunctuur op lichte

Tabel Criterialijst, met achter elk criterium het aantal punten dat wordt toegekend, als een studie eraan voldoet.

Vergelijkbaarheid der groepen	
• Homogeniteit	2
• Prestratificatie/balancing	2
• Randomisatie	4
• Tabel met relevante baseline-karakteristieken	1
• <20% uitvallers*	2
Effectmeting vergelijkbaar	
• Patiënten blind	4
• Evaluator blind	3
• Relevante follow-up	
- ≥ 3 maanden	2
- pijn	2
- algemene dagelijkse levensverrichtingen (ADL)	2
- medicijngebruik	2
Extra punten	
• Pragmatisch experiment	3
• ≥ 50 personen per groep	3
• DNIC-omzeiling (non-perforatie)**	1
• Therapie goed beschreven	2
• Presentatie met ruwe data	3
• Bijwerkingen vermeld	1
• Kwaliteit van de acupuncturist vermeld	1
Minpunten	
• Alleen p-waarden in de presentatie van de resultaten	-3

* Voor cross-over experimenten is het criterium <10% gehanteerd.

** De afkorting DNIC staat voor Diffuse Noxious Inhibitory Control.⁹ Zeer kort geformuleerd zegt de DNIC-theorie dat pijnlijke prikkels die op extrasegmentale plaatsen worden toegediend, de verwerking van reeds aanwezige pijnimpulsen op spinaal niveau beïnvloeden. In het artikel over geschiedenis en werkingsmechanismen wordt het verschil tussen DNIC en de poorttheorie van *Melzack and Wall* toegelicht.

vormen van astma zouden gemist kunnen worden, als te veel patiënten met ernstige vormen worden opgenomen in het onderzoek.

Homogeniteit wordt in een bepaalde setting (die ook beschreven dient te worden) bereikt met het gebruiken van goede in- en uitsluitcriteria. Het probleem schuilt in de term 'prognostisch relevante kenmerken'. Welke kenmerken zijn relevant voor de prognose? Hierover lopen de meningen soms uiteen. Soms ontbreekt voldoende onderzoek. Kortom, dit criterium laat zich moeilijk operationaliseren en het is met zeer grote voorzichtigheid gescoord.

Prestratificatie en *balancing* zijn technieken die tot doel hebben de onderzoeksgroepen prognostisch vergelijkbaar te maken voor bekende *confounders* (vertekende factoren). *Randomisatie* streeft hetzelfde doel na (voor onbekende en moeilijk meetbare *confounders*), maar is bij kleine aantallen patiënten als enige methode minder betrouwbaar. Idealiter wordt na het selecteren van een homogene groep patiënten geprestratificeerd, waarna de patiënten op interventie gerandomiseerd worden binnen de verschillende strata (*figuur*).

Een *tabel met relevante baseline-karakteristieken* van de patiënten (kenmerken van de patiënten vóór aanvang van de therapie) maakt het mogelijk te controleren of op deze factoren inderdaad prognostische vergelijkbaarheid is verkregen. Opmerkingen over statistische significantie horen daar niet thuis.

Wanneer tijdens het experiment veel patiënten uitvallen (= van verdere deelname afzien), kan de beoordeling van de resultaten worden bemoeilijkt, omdat de oorspronkelijke vergelijkbaarheid van de groepen kan zijn aangetast door de (selectieve?) uitval. Het optreden van veel uitvallers in de placebogroep kan betekenen dat de blinding is verbroken. Het is ook mogelijk dat teleurstelling over het effect tot uitval leidt; in dat geval zegt uitval iets over effecten (of bijwerkingen) van een therapie.

De door ons gehanteerde 20 procent-norm is arbitrair en mild. Omdat *cross-over* studies veel gevoeliger zijn voor uitval-

lers, werd daar een strengere norm gehanteerd. Onze nadruk op prognostische vergelijkbaarheid laat zich misschien als volgt begrijpelijk maken. Stel u voor, dat u scheidsrechter bent bij een wedstrijd tussen twee hardlopers. Zou u het dan niet van groot belang vinden dat de atleten vanachter dezelfde lijn starten (sintelbanen met bochten even buiten beschouwing gelaten)?

Effectmeting vergelijkbaar

Als placebo-acupunctuur wordt toegepast, is het noodzakelijk dat de patiënten niet weten of ze zich in de acupunctuurgroep of in de placebogroep bevinden. Men zegt dan dat de patiënten *geblindeerd* zijn voor de aard van de behandeling.

Bij pragmatische experimenten is dit iets moeilijker te realiseren: het onderscheid tussen acupunctuur en bijvoorbeeld fysiotherapie is (tenzij de in het *kader* besproken *double dummy*-techniek wordt toegepast) voor iedereen duidelijk. Ons is slechts één experiment bekend waarin de *double dummy*-techniek werd gebruikt. Omdat wij pragmatisch onderzoek van groot belang vinden en goed uitgevoerd pragmatisch onderzoek een kans moet krijgen om door te dringen tot de hogere regionen van de criticalijst, werden drie punten toegekend wanneer een onderzoek een pragmatisch karakter droeg. Op die manier werd de onmogelijkheid om 4 punten te scoren op het item *patiënten blind* gecompenseerd.

Meestal wordt het effect van de behandeling vastgesteld in maten waarvan de meting gevoelig is voor bewuste of onbewuste

voorkeuren. Het is dan noodzakelijk dat de evaluatie van het effect geschiedt door personen (*evaluators*) die *geblindeerd* zijn voor de aard van de behandeling. Wanneer men het overlijden van patiënten als effectmaat kiest, is dit niet nodig.

Voldoende lange *follow-up* is nodig. Tenzij anders vermeld, wordt met de duur van de follow-up de periode bedoeld die verstrijkt tussen de laatste behandeling en de laatste effectmeting. De follow-up periode die wij noodzakelijk achten, varieert per indicatiegebied. Wanneer men geïnteresseerd is het succes van acupunctuur bij mensen die zeggen met roken te willen stoppen, is een follow-up van ten minste 6 maanden aangewezen.

Zoals vermeld, werden de relevant geachte effectmaten per indicatiegebied gekozen. In de *tabel* zijn dat bijvoorbeeld het meten van pijn, en het registreren van veranderingen in ADL-vaardigheden en in het gebruik van medicamenten.

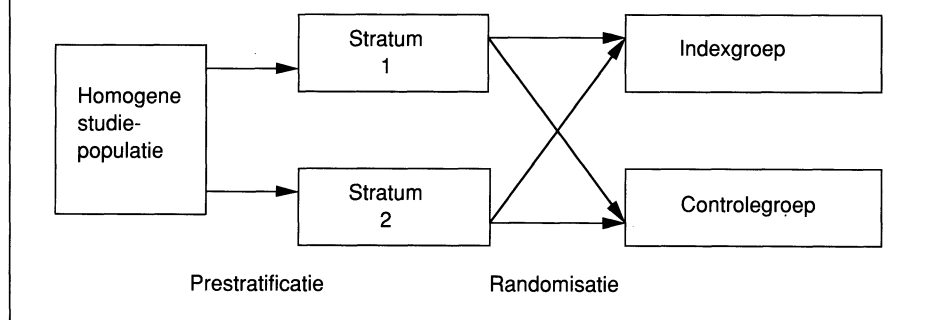
Extra punten

Wij zijn sterker geïnteresseerd in vergelijking van acupunctuur met bestaande behandelingen (*pragmatische experimenten*) dan in uitsluitend vergelijking met placebo-acupunctuur.

Om de kans op toevallsbevindingen te reduceren, is een redelijk aantal patiënten per groep gewenst. Het getal 50 is arbitrair en zeer mild gekozen.

In geval van placebo-acupunctuur waarbij de huid niet werd doorboord, gaven wij een extra punt, ondanks het feit dat de blinding van de patiënten hierdoor ern-

Figuur De toewijzing van de behandeling met behulp van prestratificatie en randomisatie.



stig werd bedreigd. Wij komen hiermee *Lewith et al.*⁹ tegemoet, die aannemelijk maken dat ook het prikken van naalden buiten acupunctuurpunten al een pijnverminderend effect kan hebben. Het begrip placebo-acupunctuur is met name op het gebied van studies naar pijn problematisch, omdat tegen de achtergrond van de DNIC-theorie (zie ook pag. 172) verondersteld moet worden dat de 'placebobehandeling' (mits enigszins pijnlijk) niet geheel zonder effect is.

Een *therapie is goed beschreven* als we kunnen lezen welke punten werden geprikt, gedurende hoeveel minuten per keer, hoe vaak per week en gedurende hoeveel weken. Andere details, zoals *Teh Ch'i* (zie *kader*) en technische specificaties bij elektro-acupunctuur, vinden we wel van belang, maar vermelding leverde geen extra punten op.

Wij vinden het van groot belang als de auteurs ons tonen wat ze hebben gemeten en niet uitsluitend gemiddelden of andere

statistisch bewerkte resultaten. Drie minpunten werden verdiend door studies die uitsluitend de uitslag van de significantietoetsen (p-waarden) gaven. Gelukkig kwam dit weinig voor.

Wanneer iets werd opgemerkt over het al dan niet *optreden van bijwerkingen*, leverde dat een punt op.

Ook als wij de indruk kregen dat de *acupuncturist* enigszins *gekwalificeerd* was, werd een punt toegekend. Het zijn vooral negatieve resultaten die aan beteke-

Een denkbeeldig experiment

Zeshonderd mannelijke polikliniekpatiënten van perifere ziekenhuizen tussen de 55 en 75 jaar oud en met een typische anamnese van claudicatio intermittens van de benen (pijn of vermoeidheid na lopen, die verdwijnt na 2-5 minuten rust) worden geselecteerd. Zij ondergaan eerst uitgebreid onderzoek, zowel omdat uitsluitcriteria moeten worden opgespoord, als om naderhand een gestratificeerde analyse te kunnen verrichten voor prognostische factoren.

Uitgesloten worden patiënten met de volgende kenmerken:

- vaatoperatie in de anamnese;
- rustpijn;
- necrose aan de voeten;
- sensibiliteitsstoornissen van de benen;
- impotent;
- aandoeningen van niet-vasculaire aard (bijvoorbeeld orthopedisch of pulmonaal) die het lopen bemoeilijken;
- eerder acupunctuur ondergaan.

Andere variabelen die worden gemeten zijn: diabetes mellitus, roken, de duur en de subjectieve ernst van de klachten, de enkel-arm-bloeddrukverhouding, de loopafstand (anamnestisch en gestandaardiseerd bij een helling van 12 procent en een loopsnelheid van 2,5 km/u), de mate van dagelijkse activiteit waarbij wordt gelopen, en de perifere bloedflow met behulp van Doppler-plethysmografie.

Er wordt geprestratificeerd op twee variabelen die waarschijnlijk een sterke prognostisch voorspellende waarde hebben: diabetes mellitus en de Doppler-uitslag (in twee categorieën).

Randomisatie (via een computerprogramma en centrale toewijzing) levert vier

groepen op. Er wordt een tabel gegeven die inzicht verschaft in hoeverre de randomisatie werkelijk voor prognostisch vergelijkbare groepen heeft gezorgd.

Alle groepen krijgen een loopprogramma (driemaal daags tot de pijntolerantiegrens lopen, gevolgd door uitrusten).

- *Groep 1* krijgt gedurende 9 maanden wekelijks een acupunctuurbehandeling (in beide benen worden de volgende punten aangeprikt: Galblaas 34, Lever 8, Nier 8. Nadat *Teh Ch'i* (een kenmerkend gevoel dat alleen zou optreden na correcte insertie van een naald in een acupunctuurpunt en dat wordt beschreven als een dof, zwaar of warm gevoel) is verkregen, blijven de stalen naalden gedurende 20 minuten zitten en worden zij elke 5 minuten kort geroosterd). Een ervaren acupuncturist voert de behandelingen uit.

Daarnaast nemen patiënten uit groep 1 negen maanden lang tweemaal daags 600 mg (= 600 IE) synthetisch d,l α -tocoferol acetaat (vitamine E) in capsules.

- *Groep 2* krijgt dezelfde acupunctuurbehandeling als groep 1. Daarnaast krijgt deze groep niet zintuiglijk te onderscheiden placebocapsules in plaats van vitamine E.

- *Groep 3* krijgt, naast de placebocapsules, een placebo-acupunctuurbehandeling die uitsluitend verschilt qua keuze van de insteekpunten en het achterwege laten van de opwekking van het *Teh Ch'i*-gevoel. In deze groep wordt eveneens een vaste combinatie van drie punten in elk been geprikt. Geen van deze punten is als acupunctuurpunt beschreven. Alle placebopunten liggen tussen 1 en 1,5 cm van de echte acupunctuurpunten uit groep 1.

- *Groep 4* wordt behandeld met vitamine E zoals in groep 1 en met placebo-acupunctuur zoals in groep 3.

Aan patiënten wordt geadviseerd geen andere therapieën te gebruiken. De placebocapsules zijn van tevoren getest op een groep vrijwilligers om er zeker te zijn dat de capsules niet met normale middelen van elkaar te onderscheiden zijn.

Omdat de procedure van *informed consent* inhoudt dat patiënten op de hoogte zijn van het gebruik van placebobehandelingen, wordt na één maand gecontroleerd of de blinding intact is; op dat moment worden nog geen therapeutische effecten verwacht, zodat een eventuele verbreking van de blinding daaraan niet kan worden geweten.

De compliantie ten aanzien van het werkelijk slikken van vitamine E wordt gemeten met behulp van een vragenlijst.

De effectmaten voor succes worden gemeten na 9 en 21 maanden en zijn: de loopafstand (gestandaardiseerd), subjectieve verbetering (gemeten via een gevalideerde vragenlijst), een vaatlabonderzoek en amputaties. Naar bijwerkingen wordt actief geïnformeerd.

Alle metingen worden verricht door een geblindeerde evaluator. Uitvallende patiënten worden actief opgespoord en gevraagd naar de redenen voor uitvallen. Analyse vindt plaats volgens het 'intention to treat'-principe.

De resultaten worden zodanig gepresenteerd dat de lezer zelf in staat is statistische bewerkingen opnieuw uit te voeren c.q. te controleren. Pas na de analyse wordt de blinding verbroken.

nis winnen wanneer men de zekerheid heeft dat de acupuncturist in de betreffende studie gekwalificeerd was.

Bij de beoordeling van *cross-over* experimenten werden enkele extra criteria gebruikt, zoals minder dan 10 procent uitvalers en een voldoende lange (meer dan 1 week) *wash out*-periode tussen de vergeleken therapieën. Het idee achter *cross-over* experimenten is dat de twee groepen patiënten na een *wash out*-periode opnieuw prognostisch vergelijkbaar zijn. Als men kiest voor een *wash out*-periode van bijvoorbeeld 2 weken, impliceert dat de verwachting dat de toegepaste behandeling niet langer dan 2 weken werkzaam is. Als men een langduriger effect veronderstelt, is een navenante *wash out*-periode noodzakelijk om een *carry over*-effect te vermijden. *Carry over*-effecten maken *cross-over* studies oninterpreteerbaar.

Sommige auteurs vonden een *wash out*-periode in zijn geheel overbodig, waardoor de resultaten uit de tweede periode van het experiment waardeloos werden (tenzij men aanneemt dat het effect van acupunctuur nog geen dag aanhoudt).

Wanneer in *cross-over* experimenten met het *Teh Ch'i*-effect wordt gewerkt, terwijl de placebo-acupunctuur op niet-acupunctuurpunten wordt toegediend, moet ervan worden uitgegaan dat de blinding van de patiënten ernstig gevaar loopt c.q. is doorbroken.¹⁰ Dit wordt veroorzaakt omdat in willekeurige huidpunten geen *Teh Ch'i* zou kunnen worden opgewekt.

Blinding acupuncturist

Men zal in de criterialijst tevergeefs zoeken naar een item 'blinding van de acupuncturist'. Er is voor zover ons bekend maar één methode waarmee men de acupuncturist effectief kan blinderen. Men spreekt dan over formule-acupunctuur waarbij de onderzoekers de acupuncturist verkeerd inlichten over de aard van de klachten van de patiënt. De acupuncturist zal dan hopelijk nietsvermoedend en met overtuiging de zijns insziens juiste behandeling instellen. In één experiment is iets dergelijks gedaan.¹¹

Andere trucs waarbij men een scheiding

aanbrengt tussen de persoon die de acupunctuurpunten c.q. de placebopunten markeert, en de persoon die de naalden insteekt, lossen het probleem slechts gedeeltelijk op; men introduceert daarmee nieuwe problemen, zoals twijfel aan de bedrevenheid van de persoon die de naalden in de vooraf gemarkeerde punten steekt, en de vraag of deze persoon werkelijk niet in staat is de juiste punten van de placebolokaties te onderscheiden.

Bij experimenten met elektro-acupunctuur kan men een technicus in een andere ruimte volgens protocol laten bepalen welke elektroden hij wel en welke hij niet op de stroombron aansluit. De markerings- en/of insteekproblemen worden hiermee echter niet opgelost.

Wij hebben de indruk dat de meeste studies zijn gedaan door onderzoekers die enthousiast tegenover acupunctuur staan. Daarom zal de vertekening die het gevolg is van een gebrekkige blinding van de acupuncturist tijdens de behandeling, leiden tot vertekening van de resultaten ten gunste van de acupunctuurbehandeling. Deze vertekening vindt dus plaats op een ander moment dan de vertekening die kan optreden bij ongeblindeerde effectmeting. Bij de interpretatie van studies die geen verschil aantonen tussen acupunctuur en placebo-acupunctuur, levert dit echter geen problemen op.

Scores

De eerste zeven items ('homogeniteit' t/m 'evaluator blind') en de laatste acht items ('pragmatisch experiment' t/m 'alleen p-waarden vermeld') komen in iedere review voor. De items onder 'relevante follow-up' zijn uiteraard aangepast aan het betreffende indicatiegebied. Soms zal de score op een bepaald criterium tussen haakjes staan. Dat is altijd een teken van onduidelijke verslaglegging of het slechts gedeeltelijk voldoen van de betreffende publikatie aan de criteria. In die gevallen is de exacte score niet precies te bepalen; de score wordt dan aangegeven in de vorm van een 'spreidingsbreedte'.

Uiteraard gaan in het proces van scoren en classificeren veel details verloren. Daarom zullen ook steeds enkele studies worden

besproken, die hoog scoorden. Dat gebeurt steeds volgens een vast stramien. In elke bespreking wordt deze volgorde gebruikt: patiëntenbeschrijving, onderzoeksdesign, beschrijving van de therapieën, effectmaten, duur van de follow-up, resultaten en kritische interpretatie.

In verband met de ruimte kunnen niet alle principes die aan modern effectiviteitsonderzoek ten grondslag liggen, uitputtend worden behandeld. Geïnteresseerden worden verwezen naar relevante literatuur.²⁻⁷

De criterialijst vormt een geïdealiseerde maatstaf en om deze iets meer inhoud te geven, is op pagina 179 een beschrijving afgedrukt van een denkbeeldig experiment dat waarschijnlijk het maximale aantal punten zou verdienen. Het gaat om een experiment dat zowel verklarend (placebo-gecontroleerd) als pragmatisch van aard is. Dat laatste houdt in dat een controlegroep een andere effectief geachte therapie krijgt. In experimenten die uitsluitend pragmatisch van aard zijn, kan men vaak niet uitsluiten dat beide therapieën op een placebo-effect berusten.

Indien het uitvalpercentage beperkt zou blijven en de blinding van de patiënten kan worden gehandhaafd, zou dit experiment 39 van de 40 punten verdienen. De DNIC-omzeiling offeren we op ten behoeve van meer zekerheid omtrent de blinding van de patiënten; wanneer men immers placebo-acupunctuur toepast waarbij de huid niet wordt geperforeerd, is de kans aanzienlijk dat patiënten dit zullen merken – zeker in het geval dat huidpunten op de voorkant van het been worden gebruikt.

Tenslotte

In deze serie zal worden geprobeerd de bezwaren van de 'gewone' reviews te vermijden. Toch is niet uitgesloten dat er nog veel ongepubliceerd materiaal bestaat (bijvoorbeeld proefschriften of studies met de auteur onwelgevallige resultaten).

De Russische literatuur werd aanvankelijk wel globaal beoordeeld, maar toen bleek dat de kwaliteit ervan stelselmatig bedroevend was, werd besloten om ten aanzien van de secundaire samenvattingen

van deze artikelen in de bestanden van Index Medicus te eisen dat de term 'randomized' of 'blind' vermeld zou staan, alvorens tot de enigszins omslachtige aanvraagprocedure en vertalen over te gaan. In de verschillende scorelijsten zal de lezer geen Russische literatuur aantreffen.

Samenvattend kan gesteld worden dat in de volgende reviews alle gecontroleerde studies kwalitatief geëvalueerd werden op een achteraf te controleren manier. Indien de lezer zelf andere gewichten aan de criteria zou willen geven, kan hij zijn eigen puntenlijst maken. Conclusies werden vooral getrokken op basis van de studies die in het schiftingsproces boven kwamen drijven.

Literatuur

- ¹ Vincent CA, Richardson PH. The evaluation of therapeutic acupuncture: concepts and methods. *Pain* 1986; 24: 1-13.
- ² Biemans MJC, Bouter LM, Knipschild PG. Effectonderzoek in de alternatieve gezondheidszorg: randomiseren en blinderen blijven integrale onderdelen. *Ned Tijdschr Integrale Geneeskde* 1987; 4: 159-68.
- ³ Ter Riet G, Kleijnen J, Knipschild PG. Over klinisch effectiviteitsonderzoek en placebo-effecten. *Integraal* 1988/89; 3: 94-101.
- ⁴ Feinstein AR. *Clinical epidemiology: the architecture of clinical research*. Philadelphia: Saunders, 1985.
- ⁵ Friedman LM, Furberg CD, DeMets DL. *Fundamentals of clinical trials*. Boston: John Wright, 1982.
- ⁶ Lubsen J, Lang de R. *Klinisch geneesmiddelenonderzoek: een praktische leidraad*. Utrecht: Bunge, 1987.
- ⁷ Bouter LM, Van Dongen MCJM. *Epidemiologisch onderzoek: opzet en interpretatie*. Utrecht: Bohn, Scheltema & Holkema, 1988.
- ⁸ Willer JC, Roby A, Le Bars D. Psychophysical and electrophysiological approaches to the pain-relieving effects of heterotopic nociceptive stimuli. *Brain* 1984; 107: 1095-1112.
- ⁹ Lewith GT, Machin D. On the evaluation of the clinical effects of acupuncture. *Pain* 1983; 16: 111-27.
- ¹⁰ Mark LC. Double-blind studies of acupuncture [Letter]. *JAMA* 1973; 225: 1532.
- ¹¹ Godfrey CM, Morgan P, Bloom B, et al. A controlled clinical trial of the theory of acupuncture in musculoskeletal pain. *J Rheumatol* 1978; 5: 612-9. ■