

# Onderzoek naar de effectiviteit van een complexe interventie in de huisartspraktijk

## Problemen bij opzet en uitvoering

H.E. VAN DER HORST

**Van der Horst HE. Onderzoek naar de effectiviteit van een complexe interventie in de huisartspraktijk. Problemen bij opzet en uitvoering. Huisarts Wet 1999;42(1):5-10.**

In dit artikel wordt een aantal methodologische problemen besproken die zich kunnen voordoen in een randomized clinical trial uitgevoerd in de huisartspraktijk, waarin het effect van een complexe interventie, gericht op gedragsverandering, wordt geëvalueerd. Als uitgangspunt is de opzet van een onderzoek genomen waarin huisartsen bij patiënten met het irritable bowel syndrome een op gedragsverandering gerichte interventie toepasten. De criteria die de Cochrane Collaboration on Effective Professional Practice Working Group heeft ontwikkeld om de methodologische kwaliteit van (randomized) clinical trials te beoordelen, zijn gebruikt als raamwerk voor de bespreking. Punten die bijzondere aandacht vereisen bij een dergelijk onderzoek, zijn: randomiseren op groepsniveau (praktijken) in plaats van op patiëntniveau; blinde of objectieve meting van de primaire uitkomstmaten; het vermijden van contaminatie van de controlegroep (= blootstellen aan delen van de interventie); het gebruiken van een adequate controlegroep; het standaardiseren van de interventie. Als laatste worden de problemen besproken die zich voordoen als, zoals vaak het geval is, tegelijkertijd de werkzaamheid en de effectiviteit van een interventie worden onderzocht. Een tweefase-opzet lijkt hiervoor de meest geschikte oplossing.

Dr. H.E. van der Horst, huisarts, Vakgroep Huisarts-, Verpleeghuis- en Sociale Geneeskunde, Vrije Universiteit, Van de Boechorststr 7, 1081 BT Amsterdam.

### Inleiding

De methodenparagraaf van een artikel dat verslag doet van de resultaten van een experimenteel onderzoek, neemt meestal een zeer beperkte ruimte in. Zelden wordt daarmee recht gedaan aan de vaak niet geringe problemen die zich voordoen bij het uitvoeren van een dergelijk onderzoek. De zogeheten beperkingen van het onderzoek komen meestal wel aan bod in de beschouwing, maar dat biedt te weinig gelegenheid om de methodologische problemen en mogelijke oplossingen uitgebreid te bespreken. Het rigoureus toepassen van de criteria van de gouden standaard – de ‘randomized clinical trial’ (RCT) – levert nogal eens problemen op bij het nagaan van de effectiviteit van een bepaald geneesmiddel. Die problemen lijken exponentieel toe te nemen wanneer het design van de RCT gebruikt wordt voor een onderzoek naar de effectiviteit van een complexe interventie. Onder een complexe interventie versta ik een interventie die uit meer dan een onderdeel bestaat en gericht is op gedragsverandering bij de patiënt, bijvoorbeeld richtlijnen voor huisartsen met betrekking tot het geven van voorlichting en adviezen voor zelfzorg.<sup>1-3</sup> In dat geval is standaardiseren van de interventie volgens protocol een belangrijke voorwaarde voor het goed uitvoeren van het onderzoek, terwijl de arts-patiëntinteractie juist kan leiden tot een afwijking van het protocol.<sup>4,5</sup> Een ander probleem is het vinden van een adequate controlegroep.<sup>6</sup>

In dit artikel beschrijf ik kort de opzet van een experimenteel onderzoek uitgevoerd onder patiënten met irritable bowel syndrome (IBS), waarin een complexe interventie is toegepast.<sup>7</sup> Daarna bespreek ik een aantal problemen die een rol spelen bij een RCT naar de effectiviteit van een complexe interventie, en geef ik suggesties voor mogelijke oplossingen ervan.

### Opzet van het IBS-onderzoek

Het onderzoek was gericht op het meten van de effectiviteit van patiëntenvoorlichting aan patiënten met IBS en het bevorderen van de zelfzorg. Daartoe is een

RCT uitgevoerd in een aantal huisartspraktijken met de volgende onderzoeksvraag: wat is de effectiviteit van het toepassen van richtlijnen voor patiëntenvoorlichting en -advisering in de huisartspraktijk bij patiënten met IBS, vergeleken met ‘gebruikelijke zorg’ op:

- zelfzorggedrag;
- het belooft van de klachten, klachtgerelateerde angst en klachtgerelateerd vermijdingsgedrag;
- de medische consumptie (aantal uitgeschreven recepten, consulten, verwijzingen)?

Een aselechte steekproef uit het Amsterdamse huisartsenbestand werd gevraagd te participeren in het onderzoek. Huisartsen die mee wilden doen, werden verdeeld in twee groepen: acht duo-praktijken en acht solopraktijken. In beide groepen werden de praktijken door het lot toegewezen aan de experimentele groep of de controlegroep.

Alle deelnemende huisartsen werd gevraagd een diagnostisch protocol te volgen en een registratieformulier in te vullen als er een patiënt met niet-acute buikklachten op het spreekuur kwam. Als volgens dit protocol de diagnose IBS werd gesteld en was voldaan aan de overige inclusiecriteria, werd de patiënt om ‘informed consent’ gevraagd.

- De huisartsen uit de *experimentele groep* kregen instructies met betrekking tot het toepassen van het diagnostisch protocol. Daarna kregen zij een drie uur durende training, waarin ze leerden de interventie toe te passen. Om de compliantie van de huisartsen met de richtlijnen na te gaan, moesten de huisartsen een checklist invullen van de consulten waarin de interventie voor de eerste keer bij een bepaalde patiënt werd toegepast. Daarnaast kregen de huisartsen het verzoek om van ten minste drie interventieconsulten een bandopname te maken. Ook van elk consult dat plaatsvond in de follow-up-periode diende de huisarts een registratieformulier in te vullen, ongeacht de klachten.

- De huisartsen uit de *controlegroep* werden alleen geïnstrueerd in het toepassen van het diagnostisch protocol en het invullen van de registratieformulieren voor de

consulten in de follow-up-periode. Zij ontvingen geen enkele instructie met betrekking tot het beleid bij IBS-patiënten. Om het contrast tussen de interventie enerzijds en de 'gebruikelijke zorg' anderzijds te kunnen vaststellen, werden de huisartsen uit de controlegroep na afloop van de gegevensverzameling ondervraagd over hun gebruikelijke zorg bij IBS.

### Methodologische problemen

De Cochrane Collaboration on Effective Professional Practice Working Group heeft zeven criteria opgesteld om de methodologische kwaliteit van (R)CTs te beoordelen.<sup>8</sup> Deze criteria zijn:

- is adequate randomisatie toegepast?
- is exclusie bias vermeden?
- is er voldoende follow-up van ingesloten patiënten?
- zijn de primaire uitkomstmaten geblindeerd gemeten?
- is er voorafgaande aan de interventie een baseline meting gedaan?
- zijn er betrouwbare primaire uitkomstmaten?
- zijn er maatregelen genomen om contaminatie te vermijden?

Drie van deze criteria – het eerste, vierde en zevende – die vaak problemen opleveren in onderzoeken in de huisartspraktijk, bespreek ik hieronder uitgebreid. Daarnaast komen drie andere problemen aan bod die buiten de Cochrane Criteria vallen.

### Randomiseren

Randomiseren wordt toegepast om selectiebias te voorkomen en om ervoor te zorgen dat eventuele storende variabelen in principe (in de praktijk werkt het toeval niet altijd mee) niet scheef verdeeld zijn over de onderzoeksgroepen. Als de primaire uitkomstmaten op patiëntniveau gemeten worden, verdient het de voorkeur om dit niveau te randomiseren. Dat is echter niet mogelijk in het geval dat de arts kennis en/of vaardigheden moet aanleren om een interventie toe te passen. Als een en dezelfde persoon beide interventies moet toepassen (bijvoorbeeld werken vol-

gens richtlijnen versus gebruikelijke zorg leveren) is het niet te vermijden dat er een contaminatie optreedt: de gebruikelijke zorg zal toch beïnvloed worden door de nieuw aangeleerde kennis.<sup>6</sup>

Om een dergelijke contaminatie te voorkomen, kan randomiseren op groepsniveau worden toegepast. Daarbij zijn er twee aandachtspunten:

- Meestal gaat het om een relatief klein aantal praktijken, waarbij de kans op een scheve verdeling van mogelijke versturende factoren (confounding) aanzienlijk is. Een scheve verdeling van een variabele die de uitkomst kan beïnvloeden, doet zich bijvoorbeeld voor als er in de interventiepraktijken veel meer mensen uit een lager sociaal-economisch milieu met slechte woonomstandigheden terechtkomen. Bekend is dat de gezondheid van deze mensen op veel terreinen minder goed is dan die van mensen uit hogere sociaal-economische milieus. Als hier geen rekening mee wordt gehouden, zou je ten onrechte kunnen concluderen dat een bepaalde interventie geen positieve effecten heeft op bepaalde aspecten van de gezondheid, omdat de interventiegroep het niet beter lijkt te doen dan de controlegroep.

Een manier om hiermee om te gaan is randomiseren in strata: er worden twee groepen gemaakt: een groep van praktijken met veel patiënten uit lagere sociale klassen en een groep met veel patiënten uit de hogere sociale klassen. Binnen deze twee groepen worden praktijken toegewezen aan hetzij de experimentele, hetzij de controlegroep. Een andere methode is om, indien mogelijk, praktijken te matchen en vervolgens het lot te laten bepalen aan welke groep elk lid van een paar wordt toegewezen.

Ook is het mogelijk achteraf na te gaan of er confounding is opgetreden door bepaalde variabelen door de verdeling ervan over de twee onderzoeksgroepen na te gaan. Mocht er sprake zijn van een scheve verdeling, dan kan daarmee rekening worden gehouden in de analyses, door bijvoorbeeld subgroepanalyses uit te voeren.

Omdat uit voorgaand onderzoek was gebleken dat de werkwijze van solistisch

werkende artsen in sommige opzichten mogelijk verschilt van de werkwijze van artsen in een groepspraktijk, is in mijn onderzoek een gestratificeerde randomisatie toegepast.<sup>9</sup>

- Als de huisartskenmerken met een mogelijk versturende invloed gelijk verdeeld zijn over de onderzoeksgroepen, is er bij randomiseren op groepsniveau nog een ander probleem dat opgelost dient te worden. Gegevens van patiënten uit dezelfde praktijk kunnen niet a priori als statistisch onafhankelijk van elkaar worden beschouwd. Het is aannemelijk dat artsen van elkaar verschillen in de wijze waarop zij met hun patiënten omgaan, en dat geldt met name bij complexe interventies. Hoewel zulke verschillen in werkwijze zoveel mogelijk uitgeschakeld dienen te worden in onderzoeken waarbij de patiënt de analyse-eenheid is, is het moeilijk om ervoor te zorgen dat bij iedere patiënt een compleet identieke interventie plaatsvindt. Daarom dient men er bij de analyse van uit te gaan dat er sprake kan zijn van een zekere mate van onderling afhankelijke uitkomstmetingen bij patiënten van dezelfde arts. Als met deze onderlinge afhankelijkheid geen rekening wordt gehouden, terwijl de analyses op patiëntniveau worden uitgevoerd, kan er een 'fallacy of the wrong level' optreden: er worden conclusies op patiëntniveau getrokken waar alleen conclusies op praktijkniveau acceptabel zijn.<sup>10</sup> Als bijvoorbeeld alleen de patiënten van slechts de helft van de huisartsen uit de experimentele groep een drastische verbetering ondergaan, waardoor de patiënten uit de experimentele groep het gemiddeld significant beter doen dan de patiënten uit de controlegroep, zou men ten onrechte kunnen concluderen dat de interventie effectief is. De enige conclusie die getrokken kan worden is dat de interventie (en waarschijnlijk nog iets anders?) effectief is, indien toegepast door huisarts x, y of z bij hun patiënten.

Door de huisarts als factor te introduceren in variantieanalyses kan nagegaan worden of en in welke mate de uitkomsten van patiënten van dezelfde huisarts onderling afhankelijk zijn. Als er sprake is van

onderlinge afhankelijkheid, dan moeten toegepaste statistische methoden worden aangepast. *Donner & Klar* bespreken verschillende procedures die gebruikt kunnen worden als de aantallen per cluster/subgroep – bijvoorbeeld een praktijk – gelijk zijn.<sup>11</sup> Het blijkt moeilijk te zijn de juiste procedure te kiezen als de aantallen per cluster verschillend zijn, zoals helaas vaak het geval is: dan blijken verschillende procedures tot nogal verschillende resultaten te leiden. Toekomstig onderzoek zal moeten uitwijzen welke keuze dan de beste is.

Een andere manier om het probleem van onderling afhankelijke uitkomsten aan te pakken, is het toepassen van multilevel-analyses. Recente ontwikkelingen in multilevel-modelling hebben statistische technieken opgeleverd met behulp waarvan individuen als onderdeel van een groep geanalyseerd kunnen worden.<sup>12</sup> Deze technieken zijn afkomstig uit de sociale wetenschappen en worden nog niet veel toegepast in de biomedische wetenschap.

Een laatste opmerking met betrekking tot randomiseren op groepsniveau betreft het berekenen van het aantal deelnemende praktijken. Het is gebruikelijk dit aantal ruwweg te berekenen door gegevens over incidentie en/of prevalentie te gebruiken, en het aantal patiënten te bepalen dat nodig is om een bepaalde power van de studie te kunnen bereiken ten aanzien van de primaire uitkomstmaat. Deze methode houdt echter geen rekening met de statistische gevolgen van het randomiseren op groepsniveau. *Diwan et al.* bespreken in hun artikel welke specifieke tabellen en power-berekeningen toegepast kunnen toegepast worden om het aantal benodigde clusters (en het aantal patiënten) vast te stellen.<sup>13</sup> Er zijn meer patiënten nodig als er op praktijkniveau gerandomiseerd wordt; daarbij geldt dat een verdubbeling van het aantal praktijken de power meer verhoogt dan een verdubbeling van het aantal patiënten.

### Primaire uitkomstmaten

De Cochrane Collaboration stelt dat de primaire uitkomstmaten hetzij geblin-

deerd, hetzij objectief – liefst geautomatiseerd met behulp van een gestandaardiseerde test – moeten worden gemeten. Schulz beschreef onlangs dat niet-dubbelblind opgezette onderzoeken grotere effecten rapporteerden dan onderzoeken met een dubbelblinde onderzoeksopzet.<sup>14</sup>

De mate van betrouwbaarheid van de uitkomstmaten die niet ‘objectief’ te meten zijn, kan worden vastgesteld door twee of meer onafhankelijke beoordelaars in te schakelen. Als de kappa van Cohen (een maat voor interbeoordelaarsbetrouwbaarheid die corrigeert voor kans) groter of gelijk is aan 0,8, wordt de betrouwbaarheid voldoende geacht.<sup>8</sup>

In het IBS-onderzoek werden sommige primaire uitkomstmaten – zoals aantal consulten, verwijzingen en recepten – verschaft door de deelnemende huisartsen die wisten in welke onderzoeksgroep ze zaten. Theoretisch is het mogelijk dat huisartsen in een of beide onderzoeksgroepen onbewust de aan te leveren gegevens manipuleren, hetzij om de onderzoeker ter wille te zijn, hetzij om hun eigen ideeën te rechtvaardigen.

Een onderzoeksassistent kan de betrouwbaarheid van de door de huisartsen geleverde gegevens controleren door ze te vergelijken met hun journaalgegevens, hoewel de volledigheid daarvan soms te wensen overlaat.<sup>15</sup> Ook kan men deze gegevens te controleren door ze te vergelijken met gegevens die de patiënt direct na de ‘event’ levert; deze methode is in praktijk echter lastig uit te voeren.

In onderzoeken als het IBS-onderzoek is zelfrapportage door patiënten een belangrijke bron van uitkomstmaten. Tot op heden is het herhaald meten van symptoomscores op Likert-schalen of visueel analoge schalen de meest betrouwbare methode voor het meten van dit soort uitkomstmaten.<sup>16</sup> De vraag is echter of feitelijk gedrag, bijvoorbeeld zelfzorgactiviteiten, ook op deze manier betrouwbaar kunnen worden gemeten. Als de vragenlijsten worden ingevoerd door een geblindeerde onderzoeksassistent of met behulp van de Optical Mark Reader, zal er geen sprake zijn van detectie-bias. Om alle mogelijkheden van fraude uit te sluiten, zou-

den de analyses geblindeerd moeten worden uitgevoerd. Omdat de onderzoeker in het algemeen uitstekend op de hoogte is van het aantal ingesloten patiënten per groep (meestal is de insluiting een belangrijk punt van zorg), is het moeilijk om aan deze eis te voldoen.

Al deze problemen zouden opgelost worden als een derde, geblindeerde persoon alle gegevens verwerkt en analyseert. Deze oplossing verhoogt echter de kosten en de complexiteit van het onderzoek.<sup>16</sup> Een compromis zou kunnen zijn een buitenstaander de analyses van de primaire uitkomstmaten te laten herhalen of controleren.

### Contaminatie vermijden

Contaminatie – het per ongeluk blootstellen van de controlegroep aan (delen van) de interventie – kan een belangrijke factor zijn in het ‘mislukken’ van een interventie-onderzoek. Dit geldt in het bijzonder als de controlegroep gevraagd wordt ‘gebruikelijke zorg’ te leveren. Om contaminatie te vermijden dient de onderzoeker aan de volgende punten aandacht te besteden.

- Een eerste vereiste om contaminatie te vermijden, is randomiseren op groepsniveau: per praktijk, per instituut, per hagro of anderszins.
- De inhoud van de interventie dient niet gepubliceerd te worden voordat de gegevensverzameling compleet is. Meestal is de onderzoeker in staat om aan deze eis te voldoen.
- De onderzoeker dient te benadrukken bij de hulpverleners die de interventie toepassen dat ze niet met anderen over de interventie praten, met name niet met hulpverleners uit de controlegroep.
- De hulpverleners uit de controlegroep moet gevraagd worden geen nascholing over het betreffende onderwerp te volgen tijdens de onderzoeksperiode.
- Tenslotte dient de onderzoeker na te gaan of de hulpverleners uit de controlegroep toevallig al elementen uit de interventie toepassen als onderdeel van hun ‘gebruikelijke zorg’.

De meest betrouwbare manier om de ‘ge-

bruikelijke zorg' en de mate van contrast daarvan met de interventie vast te stellen, is waarschijnlijk het inschakelen van simulatiepatiënten die niet als zodanig herkend worden door de hulpverlener in kwestie.<sup>17</sup> Een andere mogelijkheid is het observeren van een aantal consulten van ingesloten patiënten bij de hulpverleners uit de controlegroep, hetzij in persoon, hetzij via video/audio-opnamen.<sup>18</sup> Beide methoden vereisen nogal wat organisatie, en zijn tijd- en geldverslindend.

In mijn onderzoek heb ik de huisartsen uit de controlegroep ondervraagd aan het eind van de follow-up-periode om hun beleid bij IBS-patiënten in kaart te brengen. De huisartsen kregen een casus voorgelegd van een IBS-patiënt waarna ze aan de hand van gestructureerde vragen konden uitweiden over hun beleid. Vervolgens noemde ik de elementen van de interventie en vroeg of ze deze nooit/wel eens/ vaak/ altijd toepasten. Op deze manier kon een inschatting worden gemaakt van de grootte van het contrast tussen de interventie en de 'gebruikelijke zorg'. De huisartsen zijn pas aan het einde van de follow-up-periode ondervraagd, omdat tijdens het gesprek alle elementen van de interventie werden 'prijsgegeven'.

### Adequate controleconditie

De keuze voor een controleconditie in een onderzoek met een complexe interventie levert in het algemeen meer problemen op dan in een geneesmiddelen-RCT. *Buck & Donner* onderscheiden drie opties:

- een placebogroep;
- een alternatieve-interventiegroep;
- een onbehandelde groep.<sup>6</sup>

De eerste optie levert problemen op bij een onderzoek naar het effect van patiëntenvoorlichting: Het is onethisch om de patiënt verkeerde of nutteloze informatie te geven. De tweede mogelijkheid, een vergelijkbare, alternatieve interventie, biedt vaak geen soelaas omdat deze vaak niet voorhanden is, zeker als het gaat om patiëntenvoorlichting en/of gedragsverandering.<sup>8</sup> De derde optie, een onbehandelde groep als controle gebruiken, heeft eveneens nadelen. In de eerste plaats kan het

Hawthorne-effect hiermee niet worden uitgesloten: elke vorm van algemene aandacht voor de patiënt kan een positief gezondheidseffect produceren.<sup>8</sup> Daarnaast is het niet altijd mogelijk een onbehandelde groep te gebruiken zonder te interfereren met het hulpverleningsproces en daarmee een ethisch probleem te scheppen. Dat kan bijvoorbeeld het geval zijn als patiënten tijdens het spreekuur worden gerecruuteerd voor deelname aan een onderzoek naar klachten die nogal eens een langdurig beloop met herhaalde bemoeienis van de huisarts hebben. In dat geval is het niet mogelijk van de huisarts te vragen zich van enige vorm van beleid te onthouden.

In het IBS-onderzoek is gekozen voor de 'gebruikelijke zorg' als controleconditie, omdat het er uiteindelijk toch om gaat of de interventie meerwaarde heeft ten opzichte van de gebruikelijke zorg. Het nadeel is dat dit een niet-gestandaardiseerde controleconditie is. Als sommige vormen van 'gebruikelijke zorg' eveneens effectief zijn, kan dit leiden tot een minder groot geschat effect van de interventie bij analyses op patiëntniveau. Subgroepanalyses (vaak niet mogelijk door te kleine aantallen per subgroep) of variantieanalyses waarin de huisarts als factor is geïntroduceerd, kunnen dan uitkomst bieden.

In het IBS-onderzoek is een schatting gemaakt van de mate van contrast tussen interventie en 'gebruikelijke zorg', waaruit bleek, zoals te verwachten, dat de 'gebruikelijke zorg' nogal varieerde tussen de controlepraktijken onderling. Wel bleek er in alle gevallen sprake van een duidelijk contrast met de interventie. Variantieanalyses lieten zien dat de factor huisarts geen significante invloed op de resultaten had.

### Standaardiseren van de interventie

Een situatie waarin de interventie niet overeenkomt met wat in de ogen van de arts de beste behandeling is, kan leiden tot een belangenconflict.<sup>5,6</sup> Waarschijnlijk zal de arts afwijken van het onderzoeksprotocol als hij denkt dat dit in het belang van de patiënt is. Hierdoor komt de standaardisatie van de interventie in gevaar en daarmee

het belang van het onderzoek.

Een ander probleem is dat de arts zijn werkwijze moet veranderen en de interventie moet inpassen in de dagelijkse praktijk. De arts zal de mogelijk positieve effecten van de interventie moeten veronderstellen, omdat hij anders niet bereid is zijn werkwijze te veranderen. Hij moet ook in de gelegenheid zijn om de nieuwe 'techniek' toe te passen. Ook al gaat het om zogenaamd alledaagse klachten: de huisarts ziet ze vaak niet meer één keer per week. Als een nieuwe techniek slechts af en toe wordt toegepast, duurt het langer voordat zij in de werkwijze geïntegreerd is. Op een polikliniek is dit vaak eenvoudiger, omdat daar meestal een concentratie is van patiënten met eenzelfde soort klachten.

Dit probleem zou opgelost kunnen worden door ervoor te zorgen dat mensen met het soort klachten waar onderzoek naar gedaan wordt, worden opgeroepen om naar de praktijk te komen in een bepaalde periode, waardoor er ruime 'oefengelegenheid' ontstaat. Deze oplossing interfereert echter met de dagelijkse praktijkvoering en levert problemen op als de reden voor consultatie (bijvoorbeeld ongerustheid bij somatiserende patiënten) een belangrijke rol speelt bij de interventie. Als er echter per praktijk genoeg patiënten met bepaalde klachten zijn, zou er een kwalificatieperiode kunnen worden ingelast, waarin de huisarts bij een deel van de patiënten – die daartoe worden opgeroepen – de interventie leert toepassen. In tweede instantie volgt dan onderzoek naar de effectiviteit van de interventie. Deze kwalificatieperiode kan ook gebruikt worden om haalbaarheidsaspecten van de interventie in de dagelijkse praktijk na te gaan.

Gezien het feit dat standaardisatie van de interventie niet op voorhand aan te nemen is, dient de onderzoeker aandacht te besteden aan het vaststellen van de mate waarin de interventie conform protocol is uitgevoerd. De meest accurate methoden om vast te stellen of afwijkingen van het interventieprotocol zich hebben voorgedaan, zijn enerzijds rechtstreekse observatie van de consulten, anderzijds beoordeling van video- of audio-opnamen van

consulten. Rechtstreekse observatie van consulten is tijdrovend en meestal niet uitvoerbaar, omdat de interventieconsulten vaak niet gepland zijn. De tweede mogelijkheid is gemakkelijker uit te voeren.

In mijn onderzoek gaf echter slechts de helft van de deelnemende huisartsen uit de interventiegroep gehoor aan het verzoek om audio-opnamen te maken van twee of drie interventieconsulten. De andere huisartsen noemden verschillende redenen voor het niet inzenden van audiobandjes: bang voor mogelijk negatieve invloed op het consult, geen toestemming van patiënt, bij herhaling vergeten en technische storingen. Dit kwam overeen met wat in een eerdere onderzoek werd gerapporteerd.<sup>19</sup>

Een minder betrouwbare manier om de compliance met het interventieprotocol vast te stellen, is de deelnemende huisartsen te vragen een checklist in te vullen die de elementen van de interventie weergeeft.<sup>20</sup> In het IBS-onderzoek vulden de huisartsen voor ongeveer de helft van de interventieconsulten een checklist in. Op basis van de bandopnamen en de checklists heb ik een (helaas weinig nauwkeurige) inschatting kunnen maken van de mate waarin de huisartsen zich aan het interventieprotocol hadden gehouden.

### Effectiviteit versus werkzaamheid

Het tegelijkertijd willen vaststellen van de *werkzaamheid* (efficacy) van een complexe interventie én van de *effectiviteit* (effectiveness) van die interventie in de dagelijkse praktijk (bepaald door werkzaamheid enerzijds en haalbaarheid anderzijds) levert problemen op. In een RCT naar de effectiviteit van een medicijn is er meestal geen reden om expliciete aandacht te besteden aan de haalbaarheid, voor zover die bepaald wordt door werkwijze en praktijkvoering van de huisarts. Als het echter gaat om de effectiviteit van een complexe interventie, gebaseerd op veranderingen in de werkwijze van de huisarts, kunnen haalbaarheidsaspecten een grote rol spelen, en een zodanige negatieve invloed op de effectiviteit hebben, dat de interventie niet werkzaam lijkt te zijn. Om de werk-

zaamheid van de interventie te kunnen vaststellen, is een zorgvuldige begeleiding van de deelnemende huisartsen nodig: training, feedback en controle.

*Stephenson & Imrie* pleiten voor het uitvoeren van een RCT om de werkzaamheid van complexe, op gedragsverandering gerichte interventies vast te stellen.<sup>21</sup> Daarmee wordt echter geen informatie verkregen over de haalbaarheid en uiteindelijk de effectiviteit. Als de effectiviteit in de dagelijkse praktijk vastgesteld moet worden, dan dienen praktijkroutines zo min mogelijk doorbroken te worden, en elke huisarts zal zijn eigen weg moeten vinden in het toepassen van de interventie, hetgeen de standaardisatie van de interventie in gevaar brengt.

Ik heb geprobeerd de haalbaarheid van de interventie in de dagelijkse praktijk vast te stellen in een pilotstudie, waarvan de opbrengst tegenviel als gevolg van het feit dat ik in tijdnood kwam: binnen het tijdschema dat was opgelegd door de subsidiegever, bleek er geen ruimte te zijn om de (on)haalbaarheid goed vast te stellen. Achteraf gezien blijkt dat een aantal van de problemen die ik tegenkwam, te maken had met de onderzoeksopzet: proberen de effectiviteit van de interventie vast te stellen, terwijl noch de haalbaarheid noch de werkzaamheid vaststonden. Dit geldt met name voor het probleem van de adequaat gestandaardiseerde interventie en het probleem van de lage insluitingsgraad: slechts 15-20 procent van de in aanmerking komende patiënten werd daadwerkelijk ingesloten in het IBS-onderzoek. De belangrijkste oorzaken van deze lage insluitingsgraad waren 'vergeten eraan te denken' en tijdsdruk. Dit komt overeen met wat in andere onderzoeken werd gevonden als redenen om patiënten niet in te sluiten.<sup>22</sup>

Om de haalbaarheid van de interventie vast te stellen zou een observationele studie, gericht op de huisarts, kunnen worden uitgevoerd: de huisarts behoeft geen patiënten te werven voor het onderzoek, dient alleen bij een opeenvolgende serie patiënten de interventie uit te voeren. De onderzoeker moet de problemen die de huisartsen ondervinden, in kaart brengen

en er in samenspraak met de huisartsen een oplossing voor vinden. Om de werkzaamheid van de interventie op zichzelf vast te stellen, zou een RCT kunnen worden uitgevoerd, waarbij een getrainde huisarts de gestandaardiseerde interventie toepast bij alle ingesloten patiënten die door randomisatie zijn toegewezen aan de interventiegroep. Het identificeren van patiënten die in aanmerking komen voor het onderzoek zou in dat geval kunnen gebeuren aan de hand van een kaartenbakonderzoek. Daarmee wordt tevens selectieve in- of uitsluiting door de huisarts voorkomen. Daarmee blijft nog wel het probleem van de adequate controleconditie bestaan: is er een reële andere behandeloptie of wordt het de heterogene 'gebruikelijke zorg'.

Een alternatief zou gevonden kunnen worden in het werken met gematchte controlepatiënten: voor elke interventiepatiënt zou een controlepatiënt gevonden moeten worden, die vergelijkbaar is op een aantal relevante variabelen, en bij voorkeur uit dezelfde praktijk afkomstig is. Daarbij wordt echter het RCT-model verlaten. Ik denk dat het uitvoeren van een tweefasenonderzoek loont, als het erom gaat de effectiviteit van complexe interventies vast te stellen. Het kost weliswaar meer tijd en meer geld (hetgeen voor subsidiegevers een probleem vormt), maar het voorkomt dat er ten onrechte geconcludeerd wordt dat een interventie niet effectief is als gevolg van methodologische problemen. Bij een dergelijke tweefasenopzet is de vraag of eerst de haalbaarheid of eerst de werkzaamheid moet worden vastgesteld. Zo komt de aloude kwestie van de kip of het ei weer om de hoek kijken.

### Beschouwing

Sommige problemen die zich kunnen voordoen bij RCT's met een complexe, op gedragsverandering gebaseerde interventie in de huisartspraktijk, kunnen met enige inventiviteit worden opgelost door er in de onderzoeksopzet rekening mee te houden: randomiseren op groepsniveau (met bijbehorende sample-size-berekening, en

geëigende statistische methoden), voorkomen van contaminatie van de controlegroep, en geblindeerd vast laten stellen van uitkomstmetingen. Andere problemen die belangrijke consequenties kunnen hebben voor de validiteit van de onderzoeksresultaten, zijn: het adequaat standaardiseren van de interventie en het creëren van een adequate controleconditie. Als deze problemen onvoldoende aandacht krijgen, kunnen de resultaten van het onderzoek teleurstellend zijn, zonder dat daarmee iets wordt gezegd over de eigenlijke werkzaamheid van de interventie.

Het probleem van de tegenstrijdige belangen van het vaststellen van de werkzaamheid van een interventie op zichzelf versus het vaststellen van de effectiviteit in de dagelijkse praktijk dient ruime aandacht te krijgen bij het bepalen van de onderzoeksopzet. Een tweefasenopzet lijkt daarvoor de meest geëigende methode. Dat vereist dat ook subsidiegevers hier het belang van zullen inzien. Alleen dan zal het minder vaak voorkomen dat subsidiegeld achteraf gezien teleurstellend besteed blijkt te zijn.

### Verantwoording en dankwoord

Dit artikel is gebaseerd op hoofdstuk 5 uit mijn proefschrift, waaraan dr. F.G. Schellevis en prof.dr. J.Th.M. van Eijk een waardevolle bijdrage hebben geleverd.

### Literatuur

- 1 Pringle M, Churchill R. Randomised controlled trials in general practice. *BMJ* 1995;311:1382-3.
- 2 Silagy CA, Jewell D. Review of 39 years of randomized controlled trials in the *British Journal of General Practice*. *Br J Gen Pract* 1994;44:359-63.
- 3 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.
- 4 Hellman S, Hellman DS. Of mice but not men. Problems of the randomized clinical trial. *N Engl J Med* 1991;324:1585-9.
- 5 Schafer A. The ethics of the randomized clinical trial. *N Engl J Med* 1982;307:719-24.
- 6 Buck C, Donner A. The design of controlled experiments in the evaluation of non-therapeutic interventions. *J Chron Dis* 1982;35:531-8.
- 7 Van der Horst HE. Irritable bowel syndrome in general practice. How effective is patient education and counselling? [Dissertatie]. Amsterdam: Vrije Universiteit, 1997.
- 8 Cochrane Collaboration on Effective Professional Practice – Protocol Development Workshop. York 1995, paper.
- 9 Wijkel D. Samenwerken en verwijzen. Utrecht: Nivel, 1986.
- 10 Whiting-O'Keefe QE, Henke C, Simborg DW. Choosing the correct unit of analysis in medical care experiments. *Med Care* 1984;22:1101-14.
- 11 Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol* 1994;140:279-89.
- 12 Goldstein H, McDonald R. A general model for the analysis of multilevel data. *Psychometrika* 1988;53:455-67.
- 13 Diwan VK, Eriksson B, Sterky G, Tomson G. Randomization by group in studying the effect of drug information in primary care. *Int J Epidemiol* 1992;21:124-30.
- 14 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. *JAMA* 1995;273:408-12.
- 15 Meyboom WA. Verslaglegging van huisartsgeneeskundig handelen [Dissertatie]. Nijmegen: Katholieke Universiteit Nijmegen, 1990.
- 16 King M, Broster G, Lloyd M, Horder J. Controlled clinical trials in the evaluation of counselling in general practice. *Br J Gen Pract* 1994;44:229-32.
- 17 Rethans JJ. Does competence predict performance? [Dissertatie]. Maastricht: Rijksuniversiteit Maastricht, 1991.
- 18 Pringle M, Stewart-Evans C. Does awareness of being video recorded affect doctors' consultation behaviour? *Br J Gen Pract* 1990;40:455-8.
- 19 Faas A, Van Eijk JThM, Chavannes AW, Gubbels JW. Voorlichting ruggijn. *Huisarts Nu* 1994;8:303-9.
- 20 Abrahamson S. Evaluation in continuing medical education. *JAMA* 1968;206:625-8.
- 21 Stephenson J, Imrie J. Why do we need randomised controlled trials to assess behavioural interventions? *BMJ* 1998;316:611-3.
- 22 Peto V, Coulter A, Bond A. Factors affecting general practitioners' recruitment of patients into a prospective study. *Fam Pract* 1993;10:207-11.

### Abstract

**Van der Horst HE. Problems in the design of a clinical trial carried out in general practice. *Huisarts Wet* 1999;42(1):5-10.**

In this paper the methodological problems that might arise in a randomized clinical trial carried out in primary care and involving a complex intervention aiming at behavioural change, are discussed. The design of a clinical trial carried out among patients with irritable bowel syndrome is briefly described, in which the intervention has been applied by general practitioners. Part of the standard criteria developed by the Cochrane Collaboration on Effective Professional Practice Working Group to assess the methodological quality of (randomized) clinical trials included in their reviews are used as a framework for discussion. Problems that might arise when trying to meet these criteria are discussed, both in general and in relation to study design. Possible solutions are discussed.

Items that require special attention in the case of a trial with a complex intervention carried out by the care-providers in daily practice, are (1) random allocation at the level of the (group of) care-providers; (2) blind or objective assessment of primary outcome measurements and (3) the prevention of contamination.

We also discuss three additional problems, namely the adequacy of the control condition, the standardization of the intervention and effectiveness versus efficacy. Alternative designs for studies with complex interventions to be applied by care-providers are discussed.

**Correspondence** Dr. H.E. van der Horst MD, PhD, Department of General Practice, Free University, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands.