

J. A. KNOTTNERUS EN A. VOLOVICS

Gecontroleerd interventie-onderzoek op basis van randomisatie

Voor het medisch wetenschappelijk onderzoek wordt de 'randomized controlled trial' (RCT) beschouwd als paradigma. Het realiseren van het ideale, gerandomiseerde dubbelblinde onderzoek is echter een opgave die zelden vlekkeloos wordt volbracht en bovendien is de RCT lang niet altijd een geschikt middel. In de huisartsgeneeskunde zijn er evenwel vele vragen die wél goed te benaderen zijn met de RCT-methode. Bij veel niet-experimenteel onderzoek kunnen bovendien aspecten van de RCT-ervaring 'geleend' worden, zoals de blinding van observatoren en het streven naar vergelijkbaarheid op relevante punten. Ook voor de dagelijkse praktijk zijn sommige elementen van de RCT-methode interessant. De laatste jaren zijn in *Huisarts en Wetenschap* enkele huisartsgeneeskundige RCT's gepubliceerd. Het is belangrijk dat op dit pad wordt doorgegaan.

Inleiding

Alleen als het effect van een behandeling snel intreedt en zeer sterk is, kan dit effect zonder nauwgezet vergelijkend onderzoek worden vastgesteld. In de huisartspraktijk zijn dergelijke behandelingen zeldzaam. Dikwijls zijn er twijfels en in *Huisarts en Wetenschap* komt de behoefte aan nader onderzoek vaak naar voren, of het nu gaat om de behandeling van prolapsklachten met een pessarium, om de werkzaamheid van etsende stoffen bij wratten, om melkzuurspoelingen bij vaginitis, om de mediolaterale episiotomie ter voorkoming van totaalrupturen, om de behandeling van lage rugpijn of om de behandeling van perifere arterieel obstructief vaatlijden.¹⁻⁷

Een correcte vaststelling van het effect, met een minimum aan vertekening, is beter mogelijk in een *experimentele studie* dan in niet-experimen-

teel (observatieel) onderzoek. In het eerste geval heeft men greep op de onafhankelijke variabele (bijvoorbeeld het al of niet geven van de behandeling), de onderzoekspopulatie en de meetprocedures. Daardoor kan optimaal worden voldaan aan de volgende desiderata:

- het gewenste contrast van de te vergelijken effecten;
- de vergelijkbaarheid van de te vergelijken groepen;
- de vergelijkbaarheid van de informatieverwerking (meetprocedures) bij alle onderzochten.⁸

Bij *observatieel onderzoek* - waarvan in voorgaande afleveringen voorbeelden aan bod zijn gekomen - heeft men de situatie veel minder in de hand: in principe wordt waargenomen wat zich in de dagelijkse werkelijkheid voordoet, zij het dan op een speciale, systematische wijze. Het best haalbare hierbij is dat de kenmerken van het 'ideale' experiment zoveel mogelijk worden nagebootst.

Voor het medisch wetenschappelijk onderzoek wordt het gecontroleerde experiment op basis van randomisatie (*randomized controlled trial* - RCT) beschouwd als paradigma (*figuur 1*). Deze opzet is lang niet altijd haalbaar, maar toch zijn er in *Huisarts en Wetenschap* de laatste jaren enkele huisartsgeneeskundige RCT's gepubliceerd.⁹⁻¹¹ Het is belangrijk dat op dit pad wordt doorgegaan, omdat 'eigen' huisartsgeneeskundig effectiviteitsonderzoek nodig is.¹² De meeste klinische kennis is immers gebaseerd op verwezen patiëntenpopulaties en daarbij gaat het vaak om meer ontwikkelde en/of therapieresistente vormen van een aandoening. De prognose is in deze populaties ongunstiger en de effectiviteit van relatief eenvoudige interventies is dan niet goed meer te onderzoeken.

De RCT wordt gekenmerkt door de *randomisatie*: welke van de te vergelijken behandelingen of procedures de deelnemende patiënten ondergaan, wordt bepaald door het toeval. Hiermee wordt bereikt dat er geen systematische

verschillen optreden tussen de vergelijkingsgroepen. *Confounding by indication* (het laten bepalen van de indeling door indicaties of contra-indicaties) wordt vermeden,⁸ evenals systematische verschillen ten aanzien van andere - vaak ook onbekende - prognostische factoren. Daardoor wordt het mogelijk causale uitspraken te doen over het verband tussen behandeling en uitkomst.

Men kan randomiseren door het opgooien van een munt of een dobbelsteen. Men kan ook gebruik maken van kant-en-klare tabellen met aselechte getallen, of zulke getallen genereren via de computer. Het verdient de voorkeur bij het begin van het onderzoek een gereed randomisatieschema te hebben. 'Pseudorandomisatie', bijvoorbeeld door patiënten om en om bij een bepaalde groep in te delen, is gevaarlijk: als de arts bij één patiënt op de hoogte raakt van de indeling, kent hij daarmee de indeling van alle volgende patiënten. Deze kennis kan hem - ook onbewust - beïnvloeden bij het uitnodigen voor deelname van patiënten, of zelfs leiden tot manipulatie bij de groepsindeling.

Ongelijke groepen

Randomisatie leidt niet zonder meer tot even grote behandelingsgroepen en een (ongeveer) gelijke verdeling van patiëntgebonden en prognostische factoren over die groepen. 'Op de lange duur' - bij een voldoende groot aantal deelnemers - zal dit wel het geval zijn, maar vaak is de onderzoekspopulatie nogal klein en moet men rekening houden met verschillen.⁹ De *tabel* illustreert de kans op een *verschil in groeps-grootte* bij 'eenvoudige aselechte toewijzing' van 30 patiënten aan twee groepen. Er is maar 14 procent kans op een

Tabel De waarschijnlijkheid van diverse groeps-grootten bij randomisatie van 30 personen in twee groepen.

| Groeps-grootten | Waarschijnlijkheid |
|-----------------|--------------------|
| 15 15 | 14,4% |
| 14 16 | 27,1% |
| 13 17 | 22,3% |
| 12 18 | 16,1% |
| 11 19 | 10,2% |
| 10 20 | 5,6% |
| 9 21 | 2,7% |
| 8 22 | 1,1% |
| 7 23 (of erger) | 0,5% |

Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht.

Prof. dr. J.A. Knottnerus, hoogleraar huisartsgeneeskunde, vakgroep Huisartsgeneeskunde; A. Volovics, statisticus, vakgroep Medische Informatie en Statistiek.

Correspondentie: Prof. dr. J.A. Knottnerus.

15/15-verdeling en 20 procent kans op een wanverhouding van 11/19 of erger.

Er zijn drie mogelijkheden om ook bij kleinere onderzoekspopulaties (ongeveer) gelijke aantallen per groep te garanderen:

- opnieuw randomiseren, indien bij controle van het randomisatieschema een belangrijk verschil in de aantallen per groep wordt gevonden; ⁴⁵
- het randomiseren in aselechte 'gepermuteerde' blokken: het randomisatieschema wordt zodanig opgesplitst in kleinere blokken, dat telkens per blokje van (bijvoorbeeld 6) gerandomiseerden even grote groepen (3/3) ontstaan;
- de 'onzuivere munt'-methode: tijdens de randomisatie is de kans om in één van beide groepen terecht te komen 50 procent, tenzij een van de groepen groter is geworden dan de andere; in dat geval wordt de kans om in de andere groep te komen groter, net zolang tot weer evenwicht is bereikt.

Ongelijke verdeling van patiëntkenmerken of prognostische factoren – die ook kan optreden bij gelijke groeps grootten – is zeker ongewenst. Ten eerste gaat men bij belangrijke discrepanties twijfelen aan de uitvoering van de randomisatie en ten tweede wordt het moeilijk uit het onderzoek ondubbelzinnige conclusies te trekken zonder ingewikkelde statistische technieken te hanteren. Als men vergelijkbare groepen heeft, kan men in het algemeen volstaan met een eenvoudige presentatie van de belangrijkste uitkomsten met maximale overtuigingskracht. Belangrijke verschillen liggen weer op de loer bij kleine onderzoekspopulaties, vooral als zij weinig homogeen zijn, bijvoorbeeld wat betreft leeftijd, ernst van de ziekte en algemene toestand.

Zo'n discrepantie is ingeslopen in een huisartsgeneeskundige studie naar het vermageringseffect van HCG-injecties.¹¹ Van de 11 gerandomiseerden kwamen er 5 terecht in de placebogroep (behandeld met fysiologisch zout) en dit waren tevens de 5 zwaarste proefpersonen.

Dit probleem kan worden aangepakt door *gestratificeerde randomisatie*. Hierbij formeert men vooraf 'strata', bestaande uit (combinaties van) categorieën van patiëntkenmerken waarvan bekend is of vermoed wordt dat zij de uitkomst kunnen beïnvloeden (leeftijd,

geslacht, de huisarts die men heeft, enz.). De eigenlijke randomisatie wordt nu voor elk stratum apart uitgevoerd, bij voorkeur onder gebruikmaking van een methode (zoals de aselechte gepermuteerde blokken) die ervoor zorgt dat in beide groepen evenveel personen uit elk stratum terechtkomen. De groepen bevatten dan elk evenveel bejaarde mannen, jongere mannen, bejaarde vrouwen en jongere vrouwen, of evenveel personen uit iedere deelnemende huisartspraktijk.

De techniek van de gestratificeerde randomisatie moet kritisch gebruikt

worden. In kleine strata blijft de kans op verschillen in andere relevante, maar vaak onbekende kenmerken groot.

Bovendien is 'prestratificatie' op een groter aantal factoren arbeidsintensief en gecompliceerd, en wordt hierdoor de kans op fouten verhoogd. Voorts wordt het steeds moeilijker alle categorieën te vullen: indien men vooraf stratificeert op drie factoren, elk met drie waardenklassen, zijn er al $3^3=27$ subcategorieën. Men doet er dus goed aan zich in dit opzicht te beperken tot het strikt noodzakelijke.

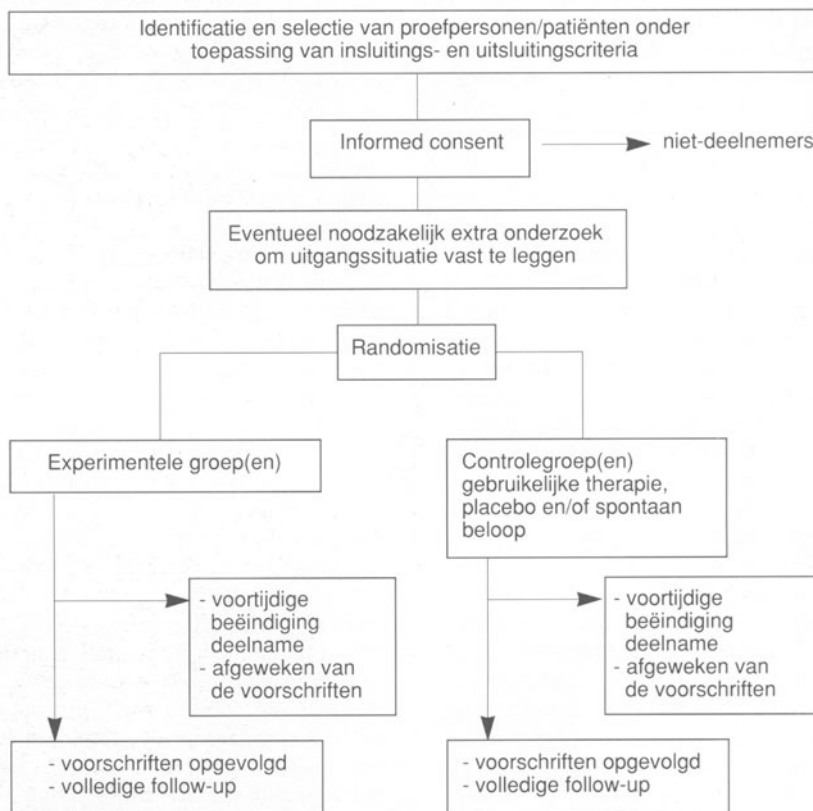
Daarnaast is het zinvol om na rando-

Figuur 1 Ontwerp van een gerandomiseerd interventie-onderzoek.

Formuleren vraagstelling, betreffende:

- toe te passen interventie(s);
- te bestuderen effect(en) of uitkomst;
- doelgroep.

Onderzoeksopzet



Analyse van de resultaten

- volgens 'intention to treat' principe;
- eventueel tussentijdse (sequentiële) analyse;
- definitieve analyse.

Interpretatie van de resultaten

misatie steeds te controleren in hoeverre inderdaad vergelijkbare groepen zijn ontstaan ten aanzien van de relevante bekende variabelen. Het is goed hierover informatie te verstrekken bij publicatie van het onderzoek, opdat de lezer zich een indruk kan vormen van het randomisatieresultaat.⁹⁻¹¹ Achteraf kunnen nog multivariabele technieken worden toegepast om te corrigeren voor ongelijk verdeelde variabelen.¹⁴

Randomisatie is mogelijk in verschillende verhoudingen. Uitgaande van twee groepen kiest men meestal voor een 1 op 1 verhouding, maar ook ongelijke verhoudingen (1 op 2, 1 op 3, etc.) zijn mogelijk. Men kan kiezen voor een ongelijke verhouding, omdat een te evalueren therapie erg duur is of door te weinig artsen wordt beheerst. Door een grotere controlegroep kan men dan het gevaar van een te kleine onderzoekspopulatie tegengaan. Bij een verhouding van 1 op 3 of nog ongelijker neemt de winst in statistische 'efficiency' echter sterk af.

Randomisatie in de praktijk

Er zijn verschillende manieren om patiënten in de huisartspraktijk te randomiseren. Men kan werken met vooraf genummerde enveloppen, die bijvoorbeeld het nummer van het te verstrekken medicijnflesje bevatten. Men kan ook kiezen voor een systeem waarbij de toewijzing voor iedere patiënt wordt verricht via een telefoontje naar een centraal *regelpunt*. Essentieel is dat steeds op centraal niveau bekend is welke patiënten meedoen, en dat de code in geval van nood (bijvoorbeeld om medische redenen) snel kan worden verbroken. Zo'n centraal overzicht is er automatisch bij de telefonische methode, maar daarbij bestaat weer het nadeel van de beperkte bereikbaarheid op bepaalde momenten.

Vraagstelling

In het algemeen geldt dat de vraagstelling van een effectiviteitsonderzoek moet aangeven:

- het *middel* waarvan de effectiviteit wordt onderzocht; dit kan een geneesmiddel zijn, een operatie, een protocol en zelfs een voorziening;
- het *effect* dat wordt onderzocht (de uitkomstparameter) en de *follow-up*

periode; het vooraf bepalen van deze follow-up periode is erg belangrijk, omdat het vaak enige tijd duurt voordat een effect waarneembaar is; anderzijds moet ook onderscheid gemaakt worden tussen de korte- en lange-termijn effectiviteit van een middel, omdat het effect van veel interventies na verloop van tijd nauwelijks meer aantoonbaar is;^{10 11 15}

- de *populatie* waarbij het effect moet worden onderzocht; deze onderzoekspopulatie wordt gedefinieerd door en getoetst aan in- en eventueel uitsluitingscriteria.

Meestal wordt de RCT toegepast voor het onderzoeken van de effectiviteit van behandelingen, maar ook diagnostica kunnen op die manier worden geëvalueerd. Het gaat dan niet zozeer om het onderscheidend vermogen van de test, maar om de invloed ervan op de prognose (waar het uiteindelijk om gaat). Voorbeelden hiervan zijn studies naar het effect van diagnostische screening (en interventie gericht) op mammacarcinoom of op risicofactoren voor harten vaatziekten.¹⁵ Ook de effectiviteit van cardiofonie kan aldus worden onderzocht.

Bij effectiviteitsonderzoek van diagnostische middelen bestaat nog de mogelijkheid om niet het uitvoeren van de test te randomiseren, maar alleen de uitslag: de behandelaar krijgt van slechts één groep de uitslag. Voordelen hiervan zijn de gemakkelijke blinding, doordat alle patiënten hetzelfde onderzoek krijgen, en het bij klinische noodzaak alsnog snel beschikbaar zijn van de betreffende informatie.

In een vorige aflevering is reeds gewezen op het belang van het juiste *contrast* in een effectvergelijking.⁸ Meting van het geïsoleerde effect van een chemische substantie vraagt om vergelijking met een placebo. *Mens* gaf bijvoorbeeld de controlegroep fysiologisch zout in plaats van HCG-injecties.¹¹ Indien men geïnteresseerd is in het totale ('bruto') effect, geeft men de controlegroep geen placebo, maar hierdoor wordt een dubbelblinde opzet onmogelijk. Meestal is het alleen relevant om de nieuwe procedure te vergelijken met het beste wat tot dan toe bekend is. Soms is het nodig de behandeling te individualiseren, zoals de dosering bij antistollings therapie. Dit stelt bijzondere eisen aan de blinding.

Cruciaal is dat men zich rekenschap geeft van het onderzochte contrast en dat vervolgens scherp formuleert. Zo werd in het onderzoek van *Mens* niet de effectiviteit van de Simeons-methode bij obesitas (dieet + HCG-injecties) onderzocht, maar alleen de additionele effectiviteit van HCG-injecties ten opzichte van fysiologisch zout. Alle proefpersonen kregen immers een dieet.¹¹

De huisartsgeneeskundige *relevantie* van de vraagstelling stelt haar eigen eisen. Zo moet men het effect van een systematisch en actief werkadvies door de huisarts op het ziekteverzuim niet vergelijken met het effect van 'nooit geven van een werkadvies'.⁹ Dit is immers een irreële optie: indien de patiënt erom vraagt, zal de huisarts vaak toch adviseren. Een betere controle-optie is daarom: 'niet actief geven van een werkadvies'. En bij het vergelijken van twee behandelingsvormen bij lage rugklachten is het niet zinvol om eerst allerlei rugpathologie uit te sluiten door middel van scanning, indien dat in de huisartspraktijk ook niet vooraf zal gebeuren.

Verder moet men besluiten of men op zoek is naar het effect van een behandeling bij maximale compliantie ('*efficacy*') of naar het effect in de praktijk, met inachtneming van eventuele non-compliantie ('*effectiviteit*'), bijvoorbeeld wegens bijwerkingen. Als in de praktijk geen dwangmaatregelen - zoals urinecontrole, pillen tellen en dagelijks laten terugkomen - zullen worden toegepast ter maximalisering van de compliantie, is kennis over effectiviteit minstens zo belangrijk als over efficacy.

Bij een vraagstelling waarbij niet een per patiënt al of niet toe te passen interventie centraal staat, maar een huisarts- of praktijkgebonden werkwijze (bijvoorbeeld een aan te leren vaardigheid), moeten niet de patiënten maar de huisartsen worden gerandomiseerd over deze (nieuwe) werkwijzen. Patiënten kunnen immers niet over huisartsen worden gerandomiseerd en tussen huisartsen bestaan a priori grote verschillen. De randomisatie-eenheid (=steekprofeenheid) is dan de arts, en het bestudeerde effect is in eerste instantie een bepaalde verandering op het niveau van de praktijk (bijvoorbeeld het verwijscijfer) en niet op het niveau van de individuele patiënten.

Wat betreft het meten van effectparameters is belangrijk, dat het onderzoek

betrekking heeft op de uitkomst waarom het in feite gaat.¹⁶ Als het bijvoorbeeld gaat om de mate waarin men weer in het dagelijks leven kan functioneren, is de uitslag van het gemakkelijker te meten elektromyogram daarvoor geen juiste maatstaf. 'An approximate answer to the right question is worth a good deal more than an exact answer to an approximate problem.'¹²

Informed consent

Het ligt voor de hand dat men toestemming vraagt alvorens men mensen opneemt in een onderzoek. Dat geldt a fortiori voor een experiment waarbij de behandeling afhankelijk is van een lotingsprocedure en, in geval van een dubbelblinde opzet, niet eens bekend wordt. Er zijn echter diverse problemen bij niet of moeilijk te blinderen (en in de huisartspraktijk veel toegepaste) interventies, zoals adviezen. Indien men hierbij expliciet om *informed consent* vraagt, verklapt men de clou, en zal bijvoorbeeld het gedrag van de controlegroep zich toch gaan verschuiven in de richting van het advies. Bovendien kan de wetenschap in welke groep men valt, op zichzelf de (rapportage van de) uitkomst gaan beïnvloeden. Het onderzoek van *Crul* naar de effectiviteit van het werkadvies door de huisarts⁹ zou om deze redenen bij het vragen van toestemming niet mogelijk zijn geweest.

Een ander probleem is dat alleen al de kennis dat men mogelijk een nepmiddel krijgt, het onderzoeksresultaat kan beïnvloeden.¹⁷ Dit speelt onder andere een rol bij bestudering van het effect van therapieën die de patiënt met volle overtuiging en inzet zou moeten ondergaan, zoals bij diverse niet-reguliere geneeswijzen de bedoeling is.

Bij onderzoek van ingrijpende therapieën schrikken patiënten, en ook artsen, vaak terug voor het idee dat hierom geloot zou moeten worden.¹⁸ Het gevoel bestaat dat het vertrouwen in de arts hierdoor wordt ondermijnd. Hiervoor is als oplossing voorgesteld dat eerst wordt geloot en vervolgens alleen toestemming wordt gevraagd aan degenen die een andere behandeling krijgen dan de gebruikelijke; van de andere patiënten wordt aangenomen dat hun situatie door het onderzoek niet verandert, en dat dus toestemming niet nodig is.¹⁹ Afgezien van de vraag of dit laatste een ethisch correcte redenering is, le-

vert dit 'Zelen-design' allerlei nieuwe methodologische problemen op. Het belangrijkste is dat de niet-deelnemers in de experimentele groep vaak relatief groot en waarschijnlijk ook selectief zal zijn ten opzichte van de controlegroep. Voorts is blinding van de patiënt – en dikwijls ook van de arts of onderzoeker – niet mogelijk.

Het is in dit licht begrijpelijk dat de regeling van informed consent sterk in de belangstelling staat.^{12, 20} Voor het patiëntgebonden huisartsgeneeskundig onderzoek gaat het daarbij vooral om het ethische spanningsveld tussen de behoefte om de individuele patiënt volledig te informeren, en de onmogelijkheid om onder die conditie de effectiviteit van veel toegepaste (en niet altijd risicoloze) interventies te onderzoeken. Waarschijnlijk zal de uitweg uit dit dilemma mede gezocht moeten worden in de wijze waarop en de mate waarin de patiënt wordt geïnformeerd.

Een absolute voorwaarde voor experimenteel onderzoek is dat er klinisch relevante *twijfel* bestaat over de effectiviteit van therapie(ën), op grond van de wetenschappelijke stand van kennis, de opvattingen onder vakgenoten, en de inzichten van de onderzoekers.

Blinding en waarnemingen

Idealiter is het onderzoek dubbelblind: patiënten noch waarnemers weten wie welke interventie krijgt. Een voorbeeld hiervan is het onderzoek van *Mens*.¹¹ Dit is vooral wenselijk indien het beloop of de waarneming beïnvloed kan worden door kennis hierover.

Soms is een 'enkelblinde' opzet het maximaal haalbare: de patiënt is wel op de hoogte van de interventie, maar de waarnemer niet. Deze situatie is goed na te bootsen in niet-experimenteel onderzoek, waarbij aldus heel wat waarnemersvertekening kan worden uitgeschakeld. Een huisartsgeneeskundig voorbeeld van deze enkelblinde opzet is de studie van *Seelen e.a.*¹⁰ De patiënt wist uiteraard of hij door de huisarts of de wijkverpleegkundige werd gecontroleerd, maar een onafhankelijk consultatiebureau verrichtte routinematig de follow-up metingen.

Blinding is niet eenvoudig en lang niet altijd mogelijk. Rekening moet worden gehouden met bijvoorbeeld kleur, reuk, smaak, vorm en toedieningswijze van het middel. Als men

in het begin al slaagt, moet men nog bedacht zijn op doorbreking van de blinding door secundaire effecten, zoals de donkere ontlasting bij ijzertherapie.

Wat de waarnemingen betreft is voorts het volgende van belang:

- De deelnemende patiëntengroepen moeten volgens hetzelfde protocol worden onderzocht en gevolgd (tevens van belang voor de blinding).
- De gebruikte meetinstrumenten moeten valide zijn, waarbij ook de inter- en intra-waarnemervariatie wordt gereduceerd door middel van ijking of training (in een volgende aflevering komen we hierop terug).
- Verschillende elementen van de waarnemingen mogen elkaar niet op ongewenste wijze beïnvloeden. Kennis over anamnestiche gegevens bijvoorbeeld kan (de resultaten van) het lichamelijke onderzoek in een bepaalde richting sturen c.q. vertekenen. In een in Zuid-Limburg lopend onderzoek met betrekking tot de diagnostiek en therapie van perifeer arterieel obstructief vaatlijden⁷ wordt om die reden het anamneseformulier door de patiënt ingevuld, het lichamelijke onderzoek volgens protocol door de huisarts verricht, en de Doppler-meting zoveel mogelijk door de praktijkassistente uitgevoerd.

De grootte van de onderzoekspopulatie

In een vorige aflevering is de essentie besproken van het vooraf schatten van de gewenste omvang van de onderzoekspopulatie. Samengevat: bij het vergelijken van het effect van twee behandelingen zijn de ingrediënten hiervan:

- het minimale verschil dat men nog voldoende relevant vindt om aan te kunnen tonen;
- het risico dat men nog aanvaardbaar vindt op het vinden van een statistisch significant verschil in effect als er in werkelijkheid geen relevant verschil is (fout van de eerste soort);
- het risico dat aanvaard wordt wat betreft het niet vinden van een significant verschil, indien er in werkelijkheid wél een relevant verschil is (fout van de tweede soort).²¹

Bij een categorische uitkomstvariabele (bijvoorbeeld wel of niet genezen) moet men voorts nog schatten hoe vaak een gunstig effect in de controlegroep wordt verwacht. In geval van een numerieke

uitkomst (zoals gewichtsverandering) is een indruk nodig van de spreiding van deze variabele in de te onderzoeken groepen. Als men bijvoorbeeld wil aantonen dat een vermageringskuur tot gemiddeld 3,5 kg meer gewichtsvermindering leidt dan een placebobehandeling, daarbij uitgaat van een standaarddeviatie van de gewichtsverandering van 3,7 kg, een fout van de eerste soort aanvaardt van 5 procent, en een fout van de tweede soort van 10 procent, dan zijn er – in geval van twee even grote groepen – minstens tweemaal 24 personen nodig. Bij een te verwachten niet-deelname van 25 procent zou dit aantal per groep 30 moeten zijn. In de studie van *Mens* is volgens deze benadering, en uitgaande van de gehanteerde premissen, impliciet een gewichtsverschil in de orde van 6 à 8 kilogram gehanteerd als nog aan te tonen verschil.

De formules waarmee deze resultaten worden bereikt, zijn weliswaar eenvoudig²², maar voor de arts-onderzoeker minder relevant dan het principe: om de gewenste grootte van de studiepopula-

tie te kunnen schatten, is *inhoudskundige* en deels subjectief gekleurde informatie nodig, die niet volgt uit statistische formules.

‘Intention to treat’

Een centraal uitgangspunt voor een juiste beoordeling van de resultaten van een RCT is de stelregel: ‘eenmaal gerandomiseerd blijft gerandomiseerd’. Alle gebeurtenissen en uitkomsten na randomisatie kunnen relevant zijn. Indien men bovendien ná randomisatie nog patiënten uit het onderzoek uitsluit, dan tast men de randomisatie aan.

Figuur 2 illustreert het belang van dit uitgangspunt. Onder degenen die behandeling A afmaakten, werd een beter resultaat bereikt dan bij degenen die B afmaakten. Bij A stopten echter veel patiënten voortijdig bij een ongunstig beloop. Bij B stopten juist veel deelnemers voortijdig vanwege het gunstige beloop. Daarom ná behandeling A in werkelijkheid eerder minder effectief dan behandeling B. Bovendien mag

men niet uitsluiten dat de patiënten die een verkeersongeval kregen, hierbij beïnvloed werden door de behandeling. Om dergelijke interpretatieproblemen te vermijden moet van alle gerandomiseerden de afloop worden vermeld. Deze afloop kan het klinische eindresultaat betreffen (wel of niet genezen; mate van bloeddrukdaling), maar ook andere eindpunten, zoals voortijdige beëindiging van de deelname. Bijgehouden moet dus worden wat er gebeurd is met degenen die de behandeling niet hebben afgemaakt, en waarom zij zijn gestopt.

Men spreekt ook wel van het ‘intention to treat’-principe. De arts wil weten wat het effect is van het *instellen* van een behandeling; het gaat niet alleen om degenen die de behandeling ook afmaken. Volgens het ‘intention to treat’-principe hadden in het onderzoek van *Crul* de 13 patiënten uit de niet-adviesgroep die op eigen verzoek toch een advies kregen,⁹ niet mogen worden uitgesloten. De mate waarin een beleid geaccepteerd en gerealiseerd wordt, moet bij dat beleid worden ingecalculleerd.

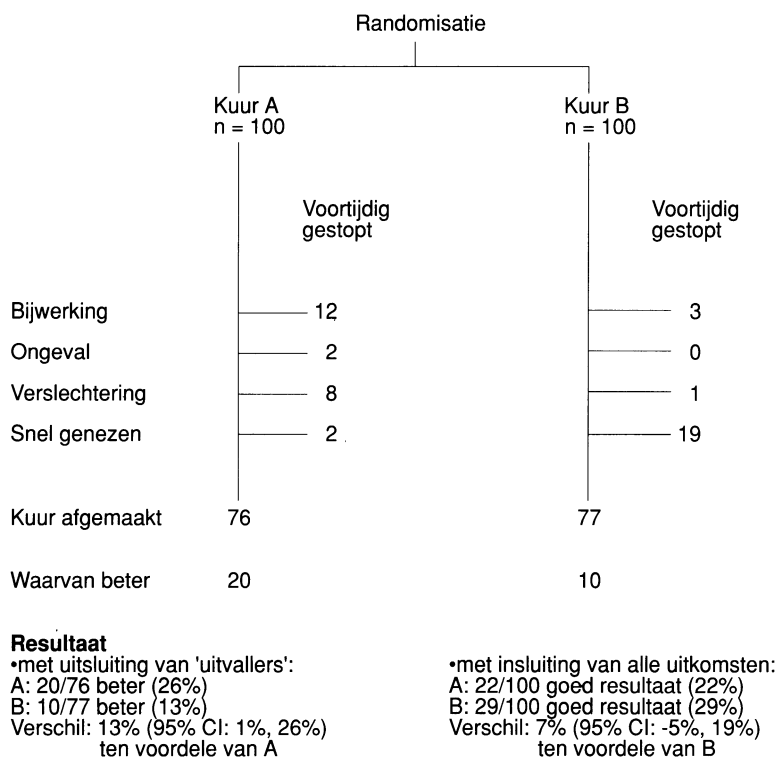
Het begrip ‘intention to treat’ sluit nauw aan bij het begrip ‘effectiviteit’ (effectiviteit van het voorschrijven van het middel met inachtneming van praktijkrelevante uitval). Als men de ‘efficacy’ van een behandeling (werkzaamheid bij maximale compliantie) wil bestuderen, zal men met allerlei kunstgrepen moeten voorkomen dat patiënten voortijdig stoppen of de voorschriften anderszins niet opvolgen.

Statistische analyse

De statistische analyse van een RCT hangt af van de gehanteerde typen variabelen (dichotoom of anderszins categorisch, dan wel numeriek). Het effect kan bijvoorbeeld beschreven worden als een verschil in genezingspercentages of een gemiddelde bloeddrukverandering. Het is hierbij voor de arts niet zozeer van belang óf er een effect is, maar vooral hoe groot dat effect is. Zoals al eerder in deze serie is uiteengezet, wordt dit beter uitgedrukt door de beschrijving van het waargenomen effect met een betrouwbaarheidsinterval, dan door het alleen vermelden van p-waarden.²¹

In tweede instantie wordt vaak nagegaan of het effect anders is in bepaalde subgroepen (bijvoorbeeld naar ge-

Figuur 2 Invloed van uitsluiting van ‘uitvallers’ op het onderzoeksresultaat (voorbeeld). Resultaten zijn gegeven met 95% betrouwbaarheidsinterval (confidence interval, CI).



slacht, leeftijd, rookgedrag e.d.), en door middel van regressie-analyse kan worden nagegaan of bepaalde additionele factoren prognostische betekenis hebben of het therapie-effect beïnvloeden.¹⁴ Ook de duur van de follow-up kan betrokken worden in de analyse. Het is immers niet alleen van belang hoe vaak patiënten bij een bepaalde behandeling genezen, maar ook hoe snel dat gebeurt. Hiervan krijgt men een indruk door het aantal bereikte genezingen te delen door het totaal aantal 'patiënten-jaren follow-up' in het kader van het onderzoek. Nog fraaier is het om de resultaten uit te zetten tegen het tijdsverloop. Zo is in *figuur 3* te zien dat de twee behandelingen M en N weliswaar hetzelfde eindresultaat hebben, maar dat M veel sneller effectief is en dus de voorkeur verdient.

Bij langdurende RCT's maakt men soms gebruik van de zogenaamde *sequentiële analyse*: van tijd tot tijd worden de resultaten geanalyseerd en naar aanleiding hiervan kan op grond van een vooraf gedefinieerde stopregel besloten worden het onderzoek voortijdig af te breken. Dat is bijvoorbeeld van belang als de resultaten al gauw zo gunstig blijken, dat het niet meer verantwoord wordt geacht de controlegroep uit te sluiten van de succesvolle therapie. Evenzo wil men een eventueel ongunstig effect van een behandeling zo vroeg mogelijk signaleren.

Behalve voordelen kent de sequentiële analyse ook beperkingen en problemen van praktische en methodologische aard. We volstaan hier met de opmerking dat naarmate er meer tussentijdse analyses worden verricht, de kans toeneemt dat toevallige fluctuaties in het beloop leiden tot het vinden van een 'fout-positief' effect. Een van de methoden om de overall kans op zo'n fout van de eerste soort in toom te houden, is het verhogen van de significantiedrempel per analyse. Bijvoorbeeld: indien men randomiseert en vijf keer een tussentijdse analyse wil uitvoeren, moet men per keer een significantieniveau van 0.016 hanteren om over het geheel niet boven de 0.050 uit te komen.

Met simpel rekenwerk komt men er niet altijd uit. Een ingrijpende therapie moet effectiever zijn dan een niet-belastende behandeling om evenveel waard te zijn, maar hoeveel effectiever? Hoeveel gun-

stiger dan het beloop bij conservatieve behandeling moet het beloop na operatieve therapie zijn om het operatierisico (later) te compenseren? Een betrekkelijk nieuw, maar voor de huisartsgeneeskunde zeer belangrijk aspect van de analyse betreft dan ook 'de kwaliteit van leven'. De enorme relevantie hiervan is evenredig met het aantal nog niet beantwoorde vragen op dit gebied.

Beschouwing

Het realiseren van het ideale, gerandomiseerde dubbelblinde onderzoek is een opgave die zelden vlekkeloos wordt volbracht. Zelfs van de gerenommeerde grote studies naar het effect van hypertensiebehandeling voldeed geen enkele aan alle eisen die gesteld zouden mogen worden.²³ Daarnaast is de RCT lang niet altijd een geschikt middel, bijvoorbeeld als de behandelingsmethoden zich zo snel ontwikkelen dat de onderzoeksresultaten verouderd zijn op het moment dat ze bekend worden. En dan hebben we het nog niet eens over de moeilijkheden in gevallen waarin zowel de te evalueren procedure als de relevante uitkomsten zeer moeilijk te meten

zijn, zoals bij de effect-evaluatie van 'methodisch werken'.²⁴

In de huisartsgeneeskunde zijn er evenwel vele vragen die wél goed te benaderen zijn met de RCT-methode. Bij veel niet-experimenteel onderzoek kunnen bovendien aspecten van de RCT-ervaring 'geleend' worden, zoals de blinding van observatoren en het streven naar vergelijkbaarheid op relevante punten. Ook voor de dagelijkse praktijk zijn sommige elementen van de RCT-methode interessant. Waarom zouden we ons in de praktijkuitoefening niet net zo druk maken over vertekening als bij wetenschappelijk onderzoek? Is een suggestieve vraagstelling tijdens de anamnese, een te gunstige afronding van de bloeddruk of een onvolledig zicht op behandelingsresultaten in de praktijk niet net zo erg?

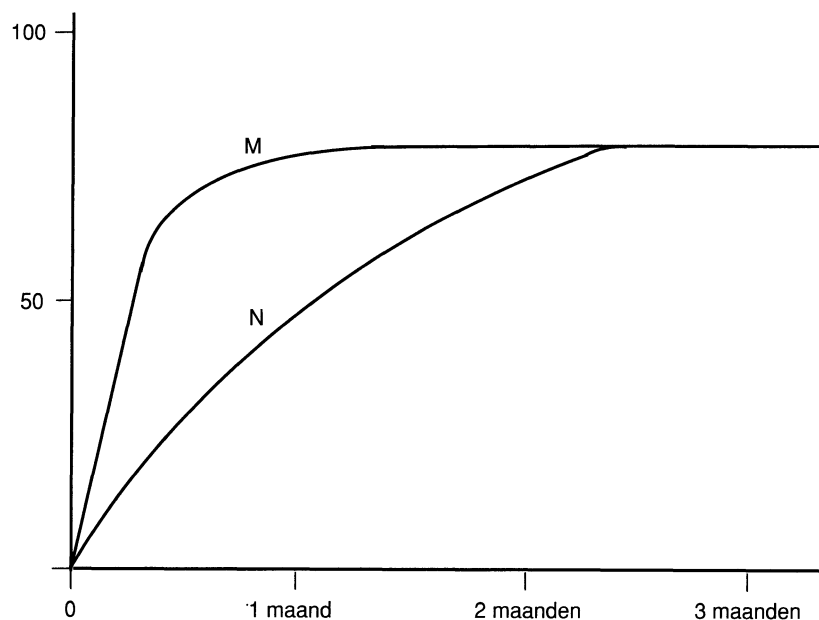
¹ Van den Bosch WJHM. Het pessarium als behandeling van prolaps van de vagina. *Huisarts Wet* 1983; 26: 411-3.

² De Haan M, Weisz FH. Wratten in de huisartspraktijk. *Huisarts Wet* 1984; 27: 162-7.

³ Boeke J, Faas A. Vaginitis in de huisartspraktijk. Een literatuuronderzoek. *Huis-*

Figuur 3 De behandelingen M en N.

Cumulatief genezingspercentage



Toelichting Beide behandelingen hebben hetzelfde eindresultaat. Omdat het effect bij M sneller intreedt, is M effectiever.

- arts Wet 1985; 28: 392-4.
- ⁴ Bergenhenegouwen NGPM. Melkzuurspoelingen bij vaginitis. Een explorerend onderzoek in een huisartspraktijk. Huisarts Wet 1985; 28: 398-9.
- ⁵ Wibaut P. De mediolaterale episiotomie en het voorkomen van totaal rupturen. Huisarts Wet 1986; 29: 198-200.
- ⁶ Rutten G, Faas A, Chavannes A. De behandeling van lage rugpijn: wat is zin-
nig? Huisarts Wet 1987; 30: 346-8.
- ⁷ Stoffers JJ, Kaiser V, Lemmens T, Knottnerus JA. Perfifeer arterieel obstructief vaatlijden: een verborgen ziektebeeld? Huisarts Wet 1988; 31: 202-6.
- ⁸ Knottnerus JA, Volovics A. Over validiteit, vergelijkbaarheid en versturende variabelen. Huisarts Wet 1988; 31: 135-40.
- ⁹ Crul BVM. De invloed van de huisarts op het kortdurend ziekteverzuim. Een vergelijkend onderzoek. Huisarts Wet 1982; 25: 102-5.
- ¹⁰ Seelen AMR, Van Zutphen WM, Diederiks JPM. Een rol voor de wijkverpleegkundige bij de interventie bij een verhoogd risico op hart- en vaatziekten. Huisarts Wet 1985; 28: 129-32.
- ¹¹ Mens J. Moeders voor snoepers? Een dubbelblind onderzoek naar het nut van injecties met humaan gonadotrofine ter ondersteuning van een vermageringsdieet. Huisarts Wet 1986; 29: 172-4.
- ¹² Knottnerus JA, de Vet HCW. Effectiviteitsonderzoek in de huisartspraktijk. Huisarts Wet 1987; 30: 49-53; 72-5; 116-9.
- ¹³ Altman DG, Gore SM. Statistics in practice. London: British Medical Association, 1982.
- ¹⁴ Knottnerus JA, Volovics A. Het analyseren van de invloed van meer dan één variabele. Huisarts Wet 1988; 31: 207-11.
- ¹⁵ Van Ree JW, Van Gerwen W, Van den Hoogen H. Interventie bij een verhoogd risico op hart- en vaatziekten. Resultaten op langere termijn van interventie bij hypercholesterolemie, adipositas en roken. Huisarts Wet 1985; 28: 21-4.
- ¹⁶ Feinstein AR. Clinimetrics. New York, London: Yale University Press, 1987.
- ¹⁷ Dahan R, Caulin C, Figea L, Kanis JA, Caulin F, Segrestaa JM. Does informed consent influence therapeutic outcome? A clinical trial of the hypnotic activity of placebo in patients admitted to hospital. Br Med J 1986; 23: 363-4.
- ¹⁸ Anonymous. Consent: how informed? Lancet 1984; ii: 1445-7.
- ¹⁹ Zelen M. A new design for randomized clinical trials. N Engl J Med 1979; 300: 1242-5.
- ²⁰ Lamberts H. Interventie-onderzoek en informed consent: een dilemma voor huisarts en patiënt? Huisarts Wet 1987; 30: 183-7.
- ²¹ Knottnerus JA, Volovics A. Statistisch toetsen. Huisarts Wet 1988; 31: 100-5.
- ²² Pocock SJ. Clinical trials, a practical ap-

proach. New York: John Wiley & Sons, 1983.

- ²³ Mangnus B. Behandeling van hypertensie: negen 'clinical trials' doorgeleefd. Huisarts Wet 1986; 29: 393-4.
- ²⁴ Sluys E. Is methodisch werken meetbaar? Huisarts Wet 1983; 26: 290-6.

Vervolg van pag. 292

- ¹ Conradi MH, Schuling J, Mulder H. Fouten in de huisartspraktijk. 1. De fouten. Huisarts Wet 1988; 31: 222-7.
- ² Conradi MH. Fouten van huisartsen. Groningen: Vakgroep Gezondheidswetenschappen, Rijksuniversiteit Groningen, 1987.
- ³ Freidson E. De medische professie: een studie van de sociologie van toegepaste kennis. Loghem: De Tijdstroom, 1981.
- ⁴ Uitspraak Medisch Tuchtcollege Amsterdam. Med Contact 1987; 42: 342-4.
- ⁵ Aghina MJ. Patiëntenrecht, een kwestie van gewicht. Assen, Amsterdam: Van Gorcum, 1978.
- ⁶ Geneeskundige Hoofdinspecteur van de Volksgezondheid. Jaarverslag 1984. Leidschendam: Ministerie van WVC, 1985.
- ⁷ Mizrahi T. Managing medical mistakes: ideology, insularity and accountability among internists-in-training. Soc Sci Med 1984; 19: 135-46.
- ⁸ Hughes EC. Mistakes at work. Can J Economics Political Sciences 1951; 17: 320-7.

Vervolg van pag. 299

- ⁵¹ Kerrebijn K. Purulente bronchitis. Afwachten of antibiotica. Vademecum 1985; 3(27).
- ⁵² Nolet HA. Het opkalefateren van mensen met een kou onder de leden. Huisarts Wet 1982; 25(suppl 6): 49-54.
- ⁵³ Anonymous. Cough and expectoration: general discussion. Eur J Resp Dis 1980; 61(suppl 110): 239-56.
- ⁵⁴ Roberts C, Imrey P, Turner J, et al. Reducing physician visits for colds through consumer education. JAMA 1983; 250: 1986-9.

Nalorfine

Meyboom vermeldt in de ideaallijst nalorfine als antidotum bij intoxicaties door opiaten.¹ Bij bespreking van deze lijst met de apotheker in ons gezondheidscentrum is gebleken dat ook nalorfine mogelijk ademdepressie geeft. Het *Informatorium medicamentorum 1988* adviseert nalorfine alleen te gebruiken als geen naloxon beschikbaar is. Nalorfine is een narcotische antagonist met agonistische werking. Ik heb daarom naloxon in mijn étui gestopt, omdat dat geen ademdepressie zou geven.

Tot nu toe heb ik het middel niet nodig gehad, zodat mijn keuze niet op ervaring berust. Is er een collega die bovenstaande kan bevestigen of ontkennen op grond van eigen ervaring?

J.B. van Nieuwkerk,
huisarts te Alkmaar

¹ Meyboom WA. Het ampullenétui – een toetsingsproject. Huisarts Wet 1988; 31: 80-8.

Naschrift

Ter voorkoming van misverstanden zij hier nogmaals vermeld dat nalorfine *niet* in de 'Ideaalijst' is opgenomen. In het tweede deel (de discussie) staat vermeld waarom niet. Door geen van de leden werd het ooit gebruikt gedurende in totaal circa vijftig praktijkjaren. Desondanks sluiten wij ook niet uit, dat elders meer ervaring kan bestaan. Het probleem van een extra middel met sterke werking bij een reeds geïntoxiceerde patiënt blijft echter hoe dan ook bestaan.

W.A. Meyboom

⁵⁵ Mookink HGA. Ziekenfondscijfers als parameter voor het handelen van huisartsen [Dissertatie]. Nijmegen. Katholieke Universiteit Nijmegen, 1986.

⁵⁶ Beek MML, Van Eijk JThM, Rutten GEHM. Hoesten: naar de dokter of niet? Een patiëntenfolder over hoesten. Huisarts Wet 1988; 31: 299-300.