

Videotoetsing van consulten van huisartsen in de eigen praktijk

Een onderzoek naar validiteit, betrouwbaarheid en haalbaarheid

P. RAM
R. GROL
J.J. RETHANS
C. VAN DER VLEUTEN

Ram P, Grol R, Rethans JJ, Van der Vleuten C. Videotoetsing van consulten van huisartsen in de eigen praktijk. Een onderzoek naar validiteit, betrouwbaarheid en haalbaarheid. *Huisarts Wet* 1999;42(10):439-45.

Doel Het ontwikkelen van een valide, betrouwbare en haalbare procedure voor videotoetsing van communicatie en medisch handelen van huisartsen in de eigen praktijk ten behoeve van kwaliteitsbevordering.

Methode In een cross-sectioneel onderzoek werden consulten van 93 huisartsen in de praktijk op video opgenomen. Op basis van tien criteria werden zestien consulten per huisarts geselecteerd. Per consult werden de arts-patiëntcommunicatie en het medisch handelen gescoord met de 'Maastrichtse Anamnese en Advies Scorelijst-Globaal'.

Resultaten De geselecteerde consulten zijn representatief voor de dagelijkse praktijk; de selectiecriteria bevorderen de standaardisatie van de procedure: solist-zijn hangt in negatieve zin samen met de score voor communicatie; de tijdsduur van de consulten beïnvloedt de score positief; de meeste huisartsen voelen zich niet beïnvloed door de video-opnamen en 90 procent van hen herkent de eigen werkstijl. De procedure is betrouwbaar na het scoren van twaalf consulten door een observator. De procedure is haalbaar: 95 procent van de video-opnamen is technisch voldoende; de acceptatie van huisartsen en patiënten is goed; de procedure kost NLG 750 per deelnemer. De Pearson correlatie-coëfficiënt tussen arts-patiëntcommunicatie en medisch handelen is 0,68, de ware coëfficiënt is 0,83.

Conclusie Videotoetsing van consulten van huisartsen in de eigen praktijk is valide en betrouwbaar, en lijkt haalbaar. De kwaliteit van arts-patiëntcommunicatie en medisch handelen hangen nauw samen.

Werkgroep Onderzoek Kwaliteit (WOK), Katholieke Universiteit Nijmegen/Universiteit Maastricht.
dr. P. Ram, huisarts, staflid NHG; prof.dr. R. Grol, hoogleraar kwaliteit van zorg; dr. J.J. Rethans, huisarts, thans Skillslab, Universiteit Maastricht; prof.dr. C. van der Vleuten, hoogleraar onderwijskunde, Universiteit Maastricht.
Correspondentie: dr. P. Ram, Universiteit Maastricht, Capaciteitsgroep Huisartsgeneeskunde, Postbus 616, 6200 MD Maastricht; e-mail Paul.Ram@HAG.Unimaas.NL.

Inleiding

Toetsing is een essentieel onderdeel van een huisartsgeneeskundig kwaliteitsbeleid.¹ Door objectief de sterke en zwakke kanten in het functioneren vast te stellen, kan men de deskundigheid van huisartsen gericht bevorderen.^{2,3} Het handelen van huisartsen kan zowel indirect als direct worden getoetst, bijvoorbeeld met behulp van kennistoetsen of prescriptie- c.q. verwijscijfers (indirect) of door middel van observaties (direct).⁴ Directe observatie van het dagelijks handelen is waarschijnlijk de meest valide methode. De feitelijk geboden zorg is immers een essentiële parameter voor kwaliteit, terwijl voldoende kennis of goede prescriptiecijfers nog geen garantie zijn voor goede zorg.⁵ Ook zijn indirecte methoden minder geschikt om de kwaliteit van de arts-patiëntcommunicatie te toetsen.

Goede communicatie is belangrijk, gezien het positieve effect op therapietrouw en angstreductie bij de patiënt.^{6,7} In welke mate de kwaliteit van de communicatie samenhangt met het medisch handelen van de huisarts, is nog niet onderzocht in de dagelijkse praktijk. Inzicht in deze samenhang kan consequenties hebben voor deskundigheidsbevordering, met name voor de mate waarin vakinhoudelijke thema's en communicatie-aspecten integraal dienen te worden aangeboden.

Het handelen van huisartsen kan getoetst worden door de huisarts tijdens zijn spreekuur te laten observeren door collega's of simulatiepatiënten, of door audio- of videoapparatuur in de praktijk te plaatsen. Deze methoden hebben elk hun voor- en nadelen. Iedere observatie kan de huisarts en de patiënt beïnvloeden.⁸

De kans op dit 'audience-effect' is het grootst als de observator in de spreekkamer aanwezig is. Bij observatie door een simulatiepatiënt speelt het effect een minder belangrijke rol, zeker als de arts niet weet dat hij geobserveerd wordt. Deze methode maakt het bovendien mogelijk de toetsing te standaardiseren: alle huisartsen krijgen dezelfde casus op identieke wijze gepresenteerd.⁹ De organisatie van deze vorm van observatie is echter complex en

de methode stelt hoge eisen aan het observerend vermogen van de simulatiepatiënt.

Het toetsen van consulten met audio-opnamen is eenvoudiger uit te voeren, maar de kwaliteit van het lichamelijk onderzoek, kleine ingrepen en non-verbale communicatie worden dan niet gemeten.

Het opnemen van consulten op video is mogelijk een goed alternatief. Video-opnamen kunnen ongestoord worden gescoord door een aantal observatoren, en ondersteunen de feedback aan de huisarts. Deze methode heeft echter ook beperkingen. Een statische camera beïnvloedt de arts mogelijk minder dan de aanwezigheid van een observator, maar geeft geen zicht op het handelen van de arts in de onderzoeksruiimte.¹⁰ Een ander probleem betreft de selectie van te observeren consulten. Dagelijkse consulten zijn steeds anders. Welke consulten komen wel en welke niet voor toetsing in aanmerking? Welke consulten zijn representatief voor het dagelijkse handelen van de huisarts en welke zijn onderling vergelijkbaar? De dagelijkse problematiek is bovendien niet gestandaardiseerd en eerder complex dan eenvoudig. Eenduidige criteria zijn moeilijk te stellen, waardoor de betrouwbaarheid van de procedure mogelijk beïnvloed wordt.¹¹

Daar komt bij dat de kwaliteit van handelen in het ene consult, bijvoorbeeld met een patiënt met blaasklachten, het handelen van die huisarts in ander soort consult matig voorspelt. Door deze 'casusspecificiteit' moeten veel consulten worden beoordeeld om tot een betrouwbare score te kunnen komen.¹² Tenslotte is bij video-opnamen de privacy van de patiënt in het geding en het is de vraag in hoeverre deze zich vrij voelt om medewerking aan de opname te weigeren.¹³

Videotoetsing is niet nieuw. In Engeland worden video-opnamen van dagelijkse consulten toegepast bij gevestigde huisartsen en huisartsen in opleiding als examenonderdeel.^{14,15} In Nederland is eerder onderzoek gedaan met behulp van videotoetsing in de praktijk.^{16,17} De validiteit, betrouwbaarheid en haalbaarheid van deze methode zijn echter nog niet volledig in kaart gebracht.

Wij hebben videotoetsing van huisartsen in de eigen praktijk onderzocht, met als doel een valide, betrouwbare en haalbare toetsing van het dagelijks handelen van huisartsen te ontwikkelen ten behoeve van deskundigheidsbevordering. Wij stelden ons daarbij de volgende vragen:

- Welke consulten komen voor toetsing in aanmerking?
- Op welke wijze kunnen deze geselecteerd worden?
- In welke mate zijn zij representatief voor de dagelijkse praktijk?
- Op welke wijze kan de methode gestandaardiseerd worden?
- In welke mate is zij betrouwbaar en haalbaar?
- Welke mate van samenhang bestaat er tussen het medisch handelen van de huisarts en diens communicatie met patiënten?

Methode

Procedure

We nodigden alle 880 huisartsen in Zuid-Nederland schriftelijk uit om aan het onderzoek deel te nemen; precies een kwart reageerde positief. Om budgettaire redenen selecteerden we 135 huisartsen at random, van wie er 35 werden getraind als observator. Achtergrondkenmerken, zoals leeftijd, geslacht en praktijkvorm, werden genoteerd om de onderzoekspopulatie te kunnen beschrijven en om de mogelijke invloed ervan op de variatie in scores te kunnen analyseren.¹⁸

Bij 93 deelnemers werd een video-observatiesysteem geïnstalleerd, bestaande uit een monitor, een recorder en twee camera's met ingebouwde microfoon (een in de spreekkamer en een in de onderzoeksruimte). Gedurende een week nam de huisarts zijn consulten integraal op door van de ene camera naar de andere te schakelen. Direct na het consult noteerde hij gegevens in een logboek, zoals leeftijd en geslacht van de patiënt, het aantal redenen van komst en de werkhypothese. De praktijkassistent informeerde de patiënt vóór het consult aan de balie over de video-opnamen, vroeg toestemming en gaf de pa-

tiënt een kaartje mee (groen = toestemming, rood = geen toestemming), dat hij bij het begin van het consult aan de huisarts overhandigde. In geval van toestemming startte deze de opname; de opname werd gewist als de patiënt achteraf zijn toestemming introk.

Validiteit

Onder de validiteit van een toets wordt verstaan: de mate waarin een toets meet wat men beoogt te meten.¹⁹ Met videotoetsing wordt beoogd het medisch en communicatief handelen van de huisarts in 'normale' consulten te meten. De te observeren consulten dienen derhalve representatief te zijn voor de dagelijkse praktijk.

Om deze representativiteit te bevorderen, werden selectiecriteria geformuleerd op basis van prevalentiecijfers van huisartsgeneeskundige aandoeningen en het takenpakket van de huisarts.²⁰⁻²⁴ Met behulp van deze criteria werden uit het logboek twintig consulten geselecteerd, zes tien voor observatie en vier als reserve. Wij stelden als norm dat in alle steekproeven aandoeningen uit minimaal acht ICPC-hoofdstukken moesten worden ingesloten (*criterium 1*); de score uit een bepaald consult voorspelt immers slechts in geringe mate de kwaliteit van handelen in een andersoortig consult.

Voor de overige criteria werd als norm gesteld dat elk criterium in minimaal 70 procent van de steekproeven geëffectueerd zou moeten zijn. Deze norm was betrekkelijk willekeurig gekozen, aangezien niet duidelijk was in hoeverre verschillende steekproefkenmerken de scores zouden beïnvloeden (zie ook onder 'Analyse').

Voor het scoren van arts-patiëntcom-

municatie en medisch handelen werd de Maastrichtse Anamnese en Advies Scorelijst-Globaal (MAAS-globaal) gebruikt.²⁵ Deze scorelijst bestaat uit drie onderdelen: communicatie per fase van het consult, algemene communicatieve vaardigheden en medisch handelen. De items zijn globaal en worden gescoord op een zevenpuntschaal (*figuur*). De MAAS-globaal is nog niet eerder in vergelijkbaar onderzoek gebruikt bij het scoren van consulten uit de dagelijkse praktijk; referentiewaarden ontbreken derhalve.

De inhoudsvaliditeit – de mate waarin de inhoud van de items een weerspiegeling vormt van het domein waarop de toets betrekking heeft – is voor het aspect arts-patiëntcommunicatie vastgesteld door experts. Deze algemeen geldende criteria zijn gebaseerd op literatuur betreffende arts-patiëntcommunicatie en in de succesvolle versies van de MAAS-Globaal geactualiseerd in een uitgebreide criteria-lijst.^{26,27} Voor medisch handelen verwijst deze criteria-lijst naar bestaande consensus binnen de beroepsgroep, richtlijnen en NHG-standaarden, omdat de geldende criteria casusspecifiek zijn.²⁸

Betrouwbaarheid

Betrouwbaarheid verwijst naar de precisie van gegevens verkregen door de toetsing. Een toets is betrouwbaar als bij herhaling bij dezelfde deelnemers dezelfde resultaten worden verkregen. Men spreekt dan van reproduceerbaarheid.

De videotoets beoogt deelnemers betrouwbaar te rangordenen op basis van verschillen tussen hun behaalde scores, de 'ware variantie'. Diverse factoren kunnen van invloed zijn op de ware variantie, zo-

Figuur MAAS-Globaal: item 8, informatie-overdracht

Algemene communicatievaardigheden							
<i>Informatie-overdracht</i>	0	1	2	3	4	5	6
Denk aan:							
- aankondigen, categoriseren							
- in kleine hoeveelheden, concrete uitleg							
- begrijpelijk taalgebruik							
- naar begrijpen vraag							
0 = afwezig; 1 = slecht; 2 = onvoldoende; 3 = twijfelachtig; 4 = voldoende; 5 = goed; 6 = uitstekend.							

als de aard en de moeilijkheidsgraad van een consult, de intra- en interbeoordelaarsbetrouwbaarheid en 'ruisfactoren' als vermoeidheid of simpelweg 'je dag niet hebben'.²⁹ Bovendien kunnen deze variantiebronnen elkaar wederzijds beïnvloeden. De ene casus zal deelnemers anders rangordenen dan de andere casus (invloed van de casus op de deelnemers) en deze verschillen tussen casus kunnen de observatoren ook beïnvloeden in hun mildheid of strengheid bij het scoren.

Door toepassing van de selectiecriteria werden de inhoud en de aard van de consulten in de steekproeven zoveel mogelijk gestandaardiseerd ten behoeve van de betrouwbaarheid en een eerlijke onderlinge vergelijking van de deelnemers. Tegelijkertijd moesten verschillen tussen de steekproefkenmerken in bepaalde mate worden geaccepteerd, omdat huisartsen nu eenmaal verschillen in werkstijl en de aard van hun patiëntenpopulatie. Op deze wijze werd er naar gestreefd zowel de validiteit als de betrouwbaarheid van de gegevens te waarborgen.

De observatoren werden in vier sessies intensief getraind in het scoren van praktijkconsulten. Daarbij werd veel aandacht besteed aan het verschil tussen observeren en interpreteren.

De zestien consulten per deelnemer werden verdeeld over drie of vier observatoren per huisarts. Iedere observator scoorde maximaal zes consulten per arts om te voorkomen dat hij te sterk beïnvloed zou worden door zijn algemene indruk van eerder gescoorde consulten, het 'halo-effect'. Voor de betrouwbaarheidsanalyse werd bij 25 procent van de deelnemers elk consult gescoord door twee observatoren.

Haalbaarheid

De acceptatie van de methode door de huisartsen is gemeten met een vragenlijst (17 vragen, vijfpuntsschaal). De haalbaarheid is afgemeten aan de technische kwaliteit van de opnamen, het aantal noodzakelijke herinstallaties, het aantal benodigde reserveconsulten en de gemaakte kosten.

Wij achtten de haalbaarheid voldoende als de meerderheid van de deelnemers zich niet beïnvloed zou voelen door de obser-

vaties, de meerderheid van de patiënten toestemming zou geven voor de opname, meer dan 90 procent van de consulten een goede beeld- en geluidskwaliteit zou hebben, en de kosten vergelijkbaar zouden zijn met andere observatiemethoden (minder dan NLG 1000 per huisarts).

Analyse

Versillen tussen de achtergrondkenmerken van de onderzoekspopulatie en die van de Nederlandse huisartspopulatie zijn geanalyseerd met de Chi-kwadraattoets en de T-toets.

Per huisarts zijn per consult en per steekproef (16 consulten) gemiddelde somscores berekend (som van alle items gedeeld door het aantal items) voor arts-patiëntcommunicatie, medisch handelen en voor beide aspecten samen (de overall-score: de som van beide scores gedeeld door 2). De mate van samenhang tussen de communicatiescores en de scores voor medisch handelen werd vastgesteld door de Pearson-correlatiecoëfficiënt te berekenen en deze te corrigeren voor de onbetrouwbaarheid van de metingen.

Per selectie criterium is nagegaan hoeveel steekproeven voldeden aan het betreffende criterium. Op deze wijze werd de mate van verschillen tussen steekproeven in kaart gebracht. Verklaring van de variantie in scores is onderzocht door een multiple regressieanalyse. De kenmerken van de huisarts – zoals leeftijd, geslacht en praktijkvorm – en de steekproefkenmerken – zoals totale duur van de zestien consulten, het aantal initiële of vervolggconsulten – waren de onafhankelijke variabelen; de gemiddelde somscores per steekproef waren de afhankelijke variabelen.

De betrouwbaarheid is gemeten met de generaliseerbaarheidsanalyse GENOVA op basis van de gemiddelde somscores per casus; met deze analyse kunnen de genoemde variantiebronnen en hun interactie worden gekwantificeerd.^{30,31} De betrouwbaarheid wordt uitgedrukt in het aantal consulten, de duur daarvan in minuten en het aantal observatoren per consult dat nodig is om tot een reproduceerbare score te komen. In deze analyse wordt rekening gehouden met de intra- en interob-

servatorbetrouwbaarheid, de aard van het consult en de interactie tussen variantiebronnen. De analyse maakt het mogelijk om op grond van gegevens verkregen uit het werkelijke aantal geobserveerde consulten (in ons onderzoek zestien) en observatoren (in ons onderzoek één observator per consult en bij 25 procent van de deelnemers twee observatoren per consult) een inschatting te maken van de betrouwbaarheid na het scoren van een hypothetisch aantal consulten door een hypothetisch aantal observatoren. Op deze wijze kan in vervolggprojecten van videotoetsing worden gekozen voor het meest voordelige design. Een waarde van $\geq 0,8$ wordt als voldoende betrouwbaar beschouwd.

Resultaten

De onderzoekspopulatie verschilde niet van de populatie Nederlandse huisartsen, met uitzondering van de gemiddelde leeftijd: de onderzoekspopulatie was gemiddeld 1,2 jaar jonger ($p < 0,05$, tabel 1). Zeven huisartsen (7%) vielen uit met als redenen: 'te bedreigend' (1), 'te druk' (4) en 'op vakantie' (2). Per deelnemer werden gemiddeld 72 consulten opgenomen (spreiding 40-95).

In alle steekproeven werden consulten uit minimaal acht ICPC-hoofdstukken ingesloten (tabel 2).

De gemiddelde somscore voor arts-patiëntcommunicatie was lager dan de somscore voor medisch handelen ($p < 0,05$, tabel 3). Op basis van de criteria is de gemiddelde score voor het medisch handelen voldoende en de gemiddelde score voor arts-patiëntcommunicatie onvoldoende. De criteria zijn echter niet onderling gewogen.

De correlatiecoëfficiënt tussen de gemiddelde scores voor arts-patiëntcommunicatie en medisch handelen was 0,68 (Pearson) en 0,83 na correctie voor de onbetrouwbaarheid van de metingen.

Het achtergrondkenmerk 'solo-huisarts' hing samen met de score voor arts-patiëntcommunicatie: solisten hadden naar verhouding een lagere score ($p < 0,05$). Geen van de overige achter-

grondkenmerken droeg bij aan het verklaren van de variantie in score, noch voor arts-patiëntcommunicatie noch voor medisch handelen. Met betrekking tot steekproefkenmerken hing alleen de totale duur van de geselecteerde consulten (gemiddeld 156,62 minuten, spreiding 100-200, SD 19,62) samen met de somscores, zowel voor communicatie als voor medisch handelen ($p < 0,01$). Het effect is positief: hoe langer de steekproef duurt, des te hoger de somscore, gegeven het selectiecriteria 'consultduur ≥ 5 en ≤ 15 minuten'.

De overall-score (gemiddelde score van het hele consult inclusief medisch handelen en arts-patiënt communicatie) was betrouwbaar, nadat acht consulten waren gescoord door twee observatoren per consult, of twaalf consulten door één observator per consult. In het eerste geval vereist dat 16 observaties, in het tweede geval 12 (tabel 4). Deze laatste aanpak is dus minder kostbaar. Voor een reproduceerbare score voor alleen de arts-patiëntcommunicatie of alleen het medisch handelen waren respectievelijk 12 en 16 consulten vereist, wat gelijk staat aan respectievelijk 24 en 16 observaties. Het blijkt dus voordeliger per consult één observator te laten scoren.

Wegens onvoldoende kwaliteit van beeld of geluid moest bij 3 procent van de deelnemers de videoapparatuur opnieuw geïnstalleerd worden en moest 5 procent van de geselecteerde consulten vervangen worden door reserveconsulten. Uiteindelijk waren alle huisartsen in staat in de spreekkamer en in de onderzoeksruimte video-opnamen te maken van voldoende technische kwaliteit.

In de schriftelijke evaluatie (respons 95%) beoordeelden de meeste huisartsen (82%) de 'informed consent'-procedure als helder, bleek 80 procent van de patiënten toestemming gegeven te hebben en vond 10 procent van de huisartsen de video-opnamen belastend voor de patiënt; dertig huisartsen (32%) voelden zich permanent beïnvloed door de video-opnamen; tien van hen schatten dat effect als positief in en twintig als negatief. Achtentwintig huisartsen voelden zich alleen tijdens de eerste consulten beïnvloed en

Tabel 1 Kenmerken van de onderzoekspopulatie (n=93) en de Nederlandse huisartsenpopulatie in 1995 (n=6549). Percentages; leeftijd in jaren

	Onderzoekspopulatie	Nederlandse huisartsen
Mannelijk geslacht	88	85
Solo-praktijk	44	51
NHG-lidmaatschap	71	69
Stadspraktijk	54	59
Huisartsopleider/docent	55	onbekend
Gemiddelde leeftijd	43	44

Tabel 2 Criteria voor de selectie van videoconsulten (16 per huisarts); per criterium het percentage steekproeven dat hieraan voldeed

1 Minimaal acht consulten uit verschillende ICPC-hoofdstukken	100
2 De eerste vijf videoconsulten zijn niet ingesloten	95
3 Acht consulten zijn te relateren aan NHG-standaarden	86
4 Representativiteit van de ICPC-hoofdstukken: D, K, L, R en S71	
5 Minimaal veertien consulten met een duur ≥ 5 en ≤ 15 minuten per consult	86
6 Leeftijdverdeling: jongste patiënt <18 jaar; oudste patiënt >65 jaar	84
7 Geslachtsverdeling patiënten: ≥ 6 en ≤ 10 vrouwelijke patiënten	72
8 Ten minste acht initiële consulten, ten minste vier vervolggconsulten	70
9 Maximaal twee consulten met >2 redenen van komst	93
10 Maximaal twee consulten met uitsluitend psychosociale problemen	94

Tabel 3 MAAS-Globaal-scores van huisartsen (n=93) in percentages: gemiddelde, standaarddeviatie en spreiding

	Gemiddelde (SD)	Spreiding
Arts-patiëntcommunicatie	41,7 (7,4)	25,8-63,9
Medisch handelen	62,0 (6,9)	47,1-84,1
Communicatief + medisch handelen	51,9 (6,6)	36,5-73,3

Tabel 4 Betrouwbaarheid: generaliseerbaarheidscoëfficiënten in samenhang met het aantal consulten en observatoren (obs)

Steekproefgrootte consulten/minuten	comm + med		communicatief		medisch		
	1 obs	2 obs	1 obs	2 obs	1 obs	2 obs	
4	48	0,58	0,68	0,55	0,65	0,52	0,63
8	96	0,74	0,81	0,71	0,79	0,68	0,78
12	144	0,81	0,87	0,79	0,85	0,76	0,84
16*	192	0,85	0,90	0,83	0,88	0,81	0,87
20	240	0,87	0,91	0,86	0,90	0,84	0,90

* Werkelijke aantal consulten (16) en observatoren (1 of 2). Alle uitkomsten zijn hoger dan 0,8, wat wil zeggen dat de scores betrouwbaar en dus reproduceerbaar zijn. Bij een geringer aantal consulten - bijvoorbeeld twaalf - zouden minimaal twee observatoren per consult nodig zijn geweest om de methode voldoende betrouwbaar te laten zijn.

Tabel 5 Kosten per huisarts in gulden

	12 consulten	16 consulten
Observatortraining*	90	90
Video-apparatuur*	45	45
Videobanden*	45	45
Installatie*	150	150
Observatie	225	300
Administratie	135	180
Selectie consulten	60	75
Totaal	750	885

* Vaste kosten.

vierendertig gaven aan zich geheel niet beïnvloed gevoeld te hebben. Afgezien van de eerste consulten, voelde 68 procent van de huisartsen zich niet beïnvloed door de video-opnamen. In de video-opnamen herkenden 85 deelnemers (91%) hun gebruikelijke werkstijl.

Het invullen van het logboek kostte 86 deelnemers minder dan één minuut per consult. De selectie van consulten uit het logboek aan de hand van de selectiecriteria kostte de onderzoeker gemiddeld een uur per deelnemer. Het logboek bleek volgens de observatoren de inhoud van de consulten voldoende weer te geven.

De materiële en personele kosten bedroegen NLG 750 per deelnemer (tabel 5).

Beschouwing

Met uitzondering van een klein verschil in gemiddelde leeftijd was de onderzoekspopulatie representatief voor de Nederlandse huisarts. De huisartsen namen vrijwillig deel aan het onderzoek. Deze bereidheid om zich te laten toetsen heeft mogelijk bijgedragen aan de acceptatie van de procedure. De geringe uitval (7%) en de positieve evaluatie zijn evenwel een stimulans om de methode te implementeren in de reguliere deskundigheidsbevordering van huisartsen. De volgende conclusies ondersteunen de haalbaarheid van de implementatie.

De methode is valide: er wordt gemeten wat wordt beoogd te meten: het feitelijk handelen van huisartsen in hun normale consulten. Door het toepassen van de selectiecriteria zijn de gescoorde consulten representatief voor het dagelijkse huisartsgeneeskundig handelen. De consulten zijn volledig opgenomen, zowel in de spreekkamer als in de onderzoeksruimte, waardoor de methode sterk wint aan inhoudsvaliditeit. Zorgvuldige bejegening van de patiënt tijdens het lichamelijke onderzoek is immers essentieel. Ook worden vaak aanvullende vragen gesteld en worden de eerste bevindingen aan de patiënt medegedeeld. Dit alles is niet zichtbaar en dus niet toetsbaar bij gebruik van een camera in de spreekkamer.

Integrale videoregistratie heeft ook beperkingen met potentieel nadelige invloeden op de validiteit. Door het schakelen was de huisarts zich mogelijk meer bewust van de observatie. Bovendien vergat men soms de apparatuur door te schakelen. Bediening door derden vermindert deze beperkingen, maar is kostbaar.¹⁰ Het ligt meer voor de hand de videoapparatuur te voorzien van een automatisch schakelsysteem. In ieder geval gaf 32 procent van de huisartsen direct na de opnamen aan zich beïnvloed gevoeld te hebben. Na het zien van de opnamen bleek dit 'audience effect' uiteindelijk beperkt: 91 procent van de huisartsen herkende zijn gebruikelijke werkstijl in de video-opnamen. De deelnemers hadden geen wezenlijk belang bij het geven van sociaal wenselijke antwoorden. Het onderzoek was gericht op het ontwikkelen van een toetsingsmethode – leerzaam maar vrijblijvend.

Het toepassen van criterium 2 (de eerste vijf consulten waren in 95 procent van de steekproeven niet ingesloten) heeft in ieder geval bijgedragen aan vermindering van het 'audience effect'. Het is bekend dat artsen zich bij observaties vooral in het begin bekeken voelen.³²

Gezien de mate van effectivering hebben de selectiecriteria de uniformiteit van de steekproeven en daardoor de standaardisatie van de procedure bevorderd. De verschillen in werkstijl tussen huisartsen blijven meetbaar, getuige de grote spreiding in de totale tijdsduur van de steekproeven. Dit pleit voor de validiteit van de procedure: de werkstijl van een huisarts en diens gemiddelde consultduur staan niet los van elkaar.^{33,34} In Engeland wordt de gemiddelde duur van het consult zelfs als een op zich zelf staand criterium voor kwaliteit beschouwd.^{35,36}

Informatie over de voorgeschiedenis van de patiënt en over eerdere consulten over dezelfde aandoening geven meer inzicht in de argumenten van de huisarts voor de keuze van een bepaald beleid in een vervolggconsult. Nu moest de observator daar vaak naar gissen, zeker als de huisarts zijn beleidsvoorstel niet expliciet toelichtte aan de patiënt. Het is echter de vraag of het inzien van deze gegevens

door de observator de validiteit van de toetsing zou doen verbeteren, daar de kwaliteit van de verslaglegging per arts en daarmee de onderbouwing van een bepaald beleid sterk wisselen.³⁷ Selectie criterium 8 (minimaal acht initiële consulten) beperkte overigens het aantal ingesloten vervolggconsulten.

De methode is betrouwbaar: na 2,5 uur observatie kan per huisarts een algemene uitspraak worden gedaan over de kwaliteit van de arts-patiëntcommunicatie en het medisch handelen. Deze betrouwbaarheid is zelfs beter dan die van kennistoetsen of vaardigheidstoetsen in gestandaardiseerde simulaties.³⁸⁻⁴⁰

Het selectie criterium 'meer dan 8 consulten met aan NHG-standaarden gerelateerde aandoeningen' is geëffectueerd bij ruim 85 procent van de deelnemers, hetgeen de observator richtlijnen bood voor het scoren van het medisch handelen.²³ Bovendien waren voor het scoren van communicatie duidelijke criteria beschreven. Daarnaast zijn de observatoren in vier sessies intensief getraind in het scoren van praktijkconsulten. Deze drie gegevens hebben mogelijk bijgedragen aan de goede betrouwbaarheid van onze procedure. Het kan echter wellicht nog beter door de selectiecriteria aan te scherpen (meer consulten met richtlijnen voor medisch handelen) en de observatoren nog intensiever te trainen.

De somscore voor arts-patiëntcommunicatie was significant lager dan de somscore voor medisch handelen. Het is moeilijk hieruit eenduidige conclusies te trekken, omdat referentiewaarden en absolute normen ontbreken. Onze feedback aan de huisarts was gebaseerd op de relatieve norm: individuele scores versus het groepsgemiddelde. Ook konden scores op item-niveau onderling vergeleken worden. Voor communicatie waren de scores voor vraagverheldering, overleg over beleid en vragen naar gevoelens relatief laag. Scores voor bejegening tijdens het lichamelijke onderzoek, ordening en empathie waren relatief hoog. Deze relatieve normen kunnen huisartsen helpen zwakke en sterke kanten van hun handelen vast te stellen.

De sterke samenhang tussen de score voor arts-patiëntcommunicatie en de score voor medisch handelen onderbouwt de functionaliteit van de arts-patiëntcommunicatie voor medisch handelen.²⁷ Omgekeerd is nader onderzoek gewenst naar de hypothese dat een huisarts die adequaat medisch handelt, beter patiëntgericht kan communiceren. In ons onderzoek scoorden de observatoren voor ieder consult zowel de communicatie als het medisch handelen. Mogelijk kan de kwaliteit van het ene aspect de observator beïnvloeden in het scoren van het andere aspect, waardoor de samenhang tussen de scores is versterkt. Tijdens de observatortraining is aan dit halo-fenomeen echter veel aandacht besteed en de criteria in deze zijn onduidelijk. Het niveau van de scores en de sterke onderlinge samenhang vragen om een evenwichtige deskundigheidsbevordering in zowel de arts-patiëntcommunicatie als vakinhoudelijke onderwerpen.

De videotoetsing in de eigen praktijk is technisch haalbaar en betaalbaar. Simulatieoetsen in een 'skills-lab' zijn aanmerkelijk duurder.⁴¹ Videoapparatuur kan per HAGRO worden aangeschaft, en observatortrainingen kunnen worden aangeboden in het kader van de reguliere deskundigheidsbevordering, hetgeen kostenbesparend zal werken. Bovendien blijkt het observeren van collegae de kwaliteit van het eigen handelen ten goede te komen.⁴² Tijdens de videotoetsing kan de huisarts ongestoord spreekuur doen, mits de 'informed consent'-procedure zorgvuldig aan de balie wordt afgehandeld.

Patiënten accepteren de methode goed, zoals blijkt uit onze evaluatie en eerder onderzoek.^{43,44} De video-opnamen dienen behandeld te worden als medisch geheim teneinde de privacy van patiënt en huisarts te waarborgen. Er dienen algemene richtlijnen gemaakt te worden voor het vragen van toestemming aan de patiënt, het toelichten van het doel van de videotoetsing, het recht tot inzage, het transport van de videobanden en het opslaan en wissen van de opnamen. In ons onderzoek zijn geen kopieën van banden gemaakt; de video-opnamen werden door de videoinstallateurs in verzegelde verpakking naar de

universiteit gebracht en per aangetekende post naar de observatoren verzonden. Al het materiaal is direct na observatie aan de huisartsen teruggestuurd. Geen band raakte verloren.

Conclusie

Onze conclusie is dat videotoetsing van huisartsen volgens de beschreven procedure valide en betrouwbaar is. De methode is haalbaar, mits huisartsen ervoor gemotiveerd zijn. Videotoetsing kan een belangrijke rol spelen in de kwaliteitsbevordering van huisartsen en huisartsen in opleiding. De methode wordt inmiddels geïmplementeerd in de huisartsopleiding in Nederland.

Dankbetuiging

Met dank aan Arnold Kester, statisticus. Universiteit Maastricht.

Literatuur

- Grol R, Wensing M, Jacobs A, Baker R, editors. Quality assurance in general practice. The state of the art in Europe. Utrecht: Nederlands Huisartsen Genootschap, 1993.
- Berwick D. Continuous improvement as an ideal in health care. *New Engl J Med* 1989;320:53-6.
- Norman GR, Davis D, Lamb S, et al. Competency assessment of primary care physicians as part of a peer review program. *JAMA*, 1993;270:1046-51.
- Rethans JJ, Westin S, Hays R. Methods for quality assessment in general practice. *Fam Pract* 1996;13:468-476.
- Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Science Educ* 1996;1:41-67.
- Feinberg J. The effect of patient-practitioner interaction on compliance: a review of the literature and application in rheumatoid arthritis. *Patient Education and Counselling* 1988;11:171-87.
- Sprecher P, Thomas E, Huebner L, et al. Effects of increased physician-patient communication on patient anxiety. *Professional Psychology: Research and Practice* 1983;14:251-5.
- Wakefield J. Direct observation. In: Neufeld VR, Norman GR, editors. *Assessing clinical competence*. New York: Springer, 1985: 51-71.
- Rethans JJ, Sturmans F, Drop MJ, Van der Vleuten CPM. Assessment of performance in actual practice of general practitioners by use of standardized patients. *Br J Gen Pract* 1991;41:97-9.
- Pringle M, Stewart-Evans C. Does awareness of being video recorded affect doctor's consultation behaviour? *Br J Gen Pract* 1990;40:45-8.
- Kane MT. The assessment of professional competence. *Evaluation & the Health Professions* 1992;15:163-82.
- Elstein A, Shulman LS, Sprafka SA. *Medical problem solving. An analysis of clinical reasoning*. Cambridge, Mass: Harvard University Press, 1978.
- Bain JE, Mackay NSD. Videotaping general practice consultations [Letter]. *BMJ* 1993;307:504.
- Campbell LM, Howie JGR, Murray TS. Use of videotaped consultations in summative assessment of trainees in general practice. *Br J Gen Pract* 1995;45:137-41.
- Fellowship by assessment-criteria FBA8. *Vale of Trent faculty*. London: Royal College of General Practitioners, 1998.
- Bensing J. *Doctor-patient communication and the quality of care [Dissertatie]*. Rotterdam: Erasmus Universiteit Rotterdam, 1991.
- Pieters HM. *De Utrechtse Consult Evaluatie Methode*. Lelystad: Meditekst, 1991.
- Caulford PG, Lamb SB, Kaigas TB, et al. Physician incompetence: specific problems and predictors. *Acad Med* 1994;69(Suppl 10):16-20.
- Dousma T, Horsten A. Tentamineren. Groningen: Wolters-Noordhoff, 1989: 120-46.
- Metsemakers JFM, Höppener P, Knottnerus JA. The Registration Network Family Practices: a computerized health information system in the Netherlands. *Br J Gen Pract* 1992;42:102-6.
- Lamberts H, Wood M, editors. *ICPC. International Classification of Primary Care*. Oxford: Oxford University Press, 1987.
- Van de Lisdonk EH, Van den Bosch WJHM, Huygen FJA, Lagro-Jansen ALM. *Ziekten in de huisartspraktijk*. Tweede druk. Utrecht: Bunge, '994.
- Lamberts H. In het huis van de huisarts. Verslag van het Transitieproject. Tweede druk. Lelystad: MediTekst, 1994.
- Springer MP, redactie. *Basistakenpakket voor de huisarts*. Utrecht: Landelijke Huisartsen Vereniging, 1983.
- Van Thiel J, Van der Vleuten CPM, Kraan H. Assessment of medical interviewing skills: generalizability of scores using successive MAAS-versions. In: Harden RM, Hart IR, Mulholland M, editors. *Approaches to the assessment of clinical competence*. Dundee: Centre for Medical Education, 1992.
- Van Thiel J, Van Dalen J, Ram P. MAAS-globaal criterialijst [Interne publicatie]. Maastricht: Universiteit Maastricht, 1995.
- Rutten GEHM, redactie. *Huisarts en patiënt: richtlijnen en uitgangspunten*. Utrecht: Nederlands Huisartsen Genootschap, 1996.
- Rutten GEHM, Thomas S, redactie. *NHG-standaarden voor de huisarts*. Utrecht: Bunge/NHG, 1993.
- Van der Vleuten CPM, Wijnen WHFW. Niets praktischer dan een goede theorie: generaliseerbaarheidstheorie als instrument voor betrouwbaarheidsstudies. *Bulletin Medisch Onderwijs*

- 1991;10:2-14.
- 30 Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. The Dependability of behavioral measurements. Theory of generalizability for scores and profiles. New York: Wiley, 1972.
 - 31 Brennan RL. Elements of generalizability theory. Iowa City, Iowa: American College Testing Publications, 1983.
 - 32 Stillman PL, Regan MB, Swanson DB, Haley HLA. Sequence effect in a multiple station examination using standardized patients. In: Hart IR, Harden RM, Des Marchais J, editors. Current developments in assessing clinical competence. Montreal, Canada: Heal Publications, 1992.
 - 33 Hays RB, Jones BF, Adkins PB, McKain PJ. Analysis of videotaped consultations to certify competence. *Med J Aust* 1990;152:609-11.
 - 34 Mokkink HGA, Schellekens CMAM, Tielens VCL. Consultduur in de huisartspraktijk. *Huisarts Wet* 1993;36:285-90.
 - 35 Howie JGR, Porter AMD, Heaney DJ, Hopton JL. Long to short consultation ratio: a proxy measure of quality of care for general practice. *Br J Gen Pract* 1991;41:48-54.
 - 36 Howie JGR, Heaney DJ, Maxwell M. Measuring quality in general practice. Occasional paper 75. London: Royal College of General Practitioners, 1997.
 - 37 Rethans JJ, Martin E, Metsemakers J. To what extent do clinical notes by general practitioners reflect actual performance? A study using simulated patients. *Br J Gen Pract* 1994;44:153-6.
 - 38 Pollemans MC. Kennistoetsing bij huisartsen [Dissertatie]. Maastricht: Rijksuniversiteit Limburg, 1994.
 - 39 Newble D, Swanson D. Psychometric characteristics of the objective structured clinical examination. *Med Educ* 1988;22:325-34.
 - 40 Norcini J, Swanson D. Factors influencing testing time, requirements for simulation-based measurements: do simulations ever yield reliable scores? *Teaching and Learning in Medicine* 1989;1:85-91.
 - 41 Reznick RK, Smee S, Baumber JS, et al. Guidelines for estimating the real cost of an Objective Structured Clinical Examination. *Acad Med* 1993;68:513-7.
 - 42 Ram P. Comprehensive assessment of general practitioners [Dissertatie]. Maastricht: Universiteit Maastricht, 1998.
 - 43 Shafir MS. Patient consent to observation. *Can Fam Physician* 1995;41:1367-72.
 - 44 Campbell LM, Sullivan F, Murray TS. Videotaping of general practice consultations: effect on patient satisfaction. *BMJ* 1995;311:236. ■

Abstract

Ram P, Grol R, Rethans JJ, Van der Vleuten C. Assessment of communicative and medical performance of general practitioners in daily practice. Validity, reliability and feasibility of video observation. *Huisarts Wet* 1999;42(10): 439-45.

Objectives To develop a video assessment method for general practitioners (GPs) by analysing issues of validity, reliability and feasibility of observation of videotaped regular consultations. To determine the correlation between scores on medical performance and scores on communication with patients.

Methods In a cross-sectional study consultations of 93 GPs were video recorded in the practice during one week. The GPs registered consultation and patient data in a logbook; 16 consultations per GP were selected using preset criteria. The quality of communicative and medical performance of these consultations was assessed by GP observers with a validated instrument. The validity of the procedure was evaluated by checking the content of each GP's sample using specific sample criteria. Selection bias was estimated by multiple regression analysis, with sample characteristics as independent variables and scores on communication and medical performance as dependent variables. The influence of observation on GPs and patients was assessed by a questionnaire. Generalizability theory was used to estimate reliability. Feasibility was assessed by conducting a questionnaire, by keeping accounts,

and by checking the technical quality of the videotaped consultations.

Pearson's correlation coefficients were calculated between scores on communication with patients and scores on medical performance, which were disattenuated for the unreliability of instruments.

Results The domain of general practice was well covered in the samples; content validity was satisfactory. As regards the sample characteristics, only the total duration of consultations appeared to correlate significantly with both the score on communication and the score on medical performance. As regards GPs' characteristics, working as a sole-practitioner appeared to correlate negatively with the score on communication. A majority (71%) of GPs reported not being influenced by the observation, except in the first cases, and recognizing their usual daily performance in the videotaped consultations. An acceptable level of reliability was reached after 2.5 hours observation, i.e. 12 cases by a single observer. The method was well accepted by both GPs and patients. The costs were GBP 250 per GP. The true correlation between medical performance and communication with patients is 0.8.

Conclusion Video assessment of GPs in daily practice according to the procedures described is a valid and reliable method, one which is useful for education and quality improvement. There is a trade off between feasibility on the one hand and validity, reliability and credibility on the other. Compared to investments in observation methods in standardized settings, the costs of video observation of GPs' actual performance are acceptable.

There is a high correlation between the quality of medical performance and the quality of communication with patients. Both aspects should be implemented into continuing medical education.

Correspondence P. Ram, MD, PhD, Maastricht University, Department of General Practice, PO Box 616, 6200 MD Maastricht, The Netherlands; e-mail Paul.Ram@HAG.Unimaas.NL.