

Integrale toetsing: de voorspellende waarde van een simulatiespreekuur en kennistoetsen voor de kwaliteit van het dagelijks handelen

PAUL RAM
JAN-JOOST RETHANS
RICHARD GROL
ET AL.

Ram P, Rethans J-J, Grol R, Van der Vleuten C, Hobma S, Aretz K. Integrale toetsing: de voorspellende waarde van een simulatiespreekuur en kennistoetsen voor de kwaliteit van het dagelijks handelen. *Huisarts Wet* 2000;43(3):103-10.

Doel Het vaststellen van de meetkundige kwaliteiten van een simulatietoets en drie kennistoetsen; selecteren van de beste methoden voor kwaliteitsverbetering in de dagelijkse praktijk.

Methode Screening en zo nodig aanpassing van de toetsinstrumenten op validiteit, betrouwbaarheid en haalbaarheid; samenstelling van een integrale toets, bestaande uit een algemeen medische kennistoets, een kennistoets over medisch-technische vaardigheden, een kennistoets over arts-patiëntcommunicatie, en twee toetsmethoden voor observatie van echte en simulatieconsulten. De drie kennistoetsen werden thuis gemaakt. De 100 deelnemers werden vervolgens ingedeeld in twee vergelijkbare groepen. Groep 1 werd eerst geobserveerd in een simulatiespreekuur en daarna in de eigen praktijk; groep 2 deed hetzelfde in omgekeerde volgorde, dit om een eventueel volgorde-effect vast te stellen. In beide settings werden de consulten op video opgenomen en door observatoren gescoord op arts-patiëntcommunicatie en medisch handelen.

Resultaten Alleen het simulatiespreekuur bleek gevoelig voor een volgorde-effect. Observatie van praktijkconsulten was meer valide en beter haalbaar dan observatie van simulatieconsulten; de betrouwbaarheid van beide methoden was vergelijkbaar. Kennistoetsen voorspelden de kwaliteit van het dagelijks handelen beter dan het simulatiespreekuur. De voorspellende waarde was echter te beperkt om observatie van echte consulten te kunnen vervangen.

Conclusie Integrale toetsing van arts-patiëntcommunicatie en medisch handelen impliceert observatie van consulten in de praktijk. Gelet op de goede toepasbaarheid zijn kennistoetsen second best.

Werkgroep Onderzoek Kwaliteit (WOK), Katholieke Universiteit Nijmegen/Universiteit Maastricht: dr. P. Ram, huisarts, stafid NHG; dr. J.-J. Rethans, huisarts, thans Skillslab Universiteit Maastricht; prof. dr. R. Grol, hoogleraar kwaliteit van zorg; S. Hobma, huisarts-onderzoeker; drs. K. Aretz, algemeen research-assistente. prof. dr. C.P.M. 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

Een systematisch kwaliteitsbeleid, zoals dat wordt voorgestaan door LHV en NHG, roept de vraag op welke onderwerpen moeten worden getoetst, en op welk niveau en met welke methoden dat moet worden gedaan.^{1,2} Medische of communicatieve aspecten kunnen worden gemeten op het niveau van kennis, vaardigheden of feitelijk handelen, met bijvoorbeeld kennistoetsen, observaties in een simulatietoevoeging of toetsing in de dagelijkse praktijk. Toetsing in de dagelijkse praktijk heeft de voorkeur, omdat bij kwaliteitsverbetering de feitelijk geleverde zorg centraal staat.³⁻⁵

Toetsing op dit niveau heeft echter een aantal beperkingen. De praktijk van alledag is grillig en niet gestandaardiseerd, wat een betrouwbare vergelijking van huisartsen bemoeilijkt.⁶ Bovendien moeten veel consulten worden geobserveerd om tot een algemeen oordeel te kunnen komen, omdat een score voor een bepaalde casus de score voor een ander soort casus slecht voorspelt. Deze 'casusspecificiteit' maakt alle observatietoetsen kostbaar.⁷

Kennistoetsen zijn minder bewerkelijk, op grote schaal toepasbaar en gestandaardiseerd. Zij dienen bij voorkeur patiëntgericht te zijn, realistische praktijksituaties te weerspiegelen en voldoende klinische kernhandelingen ('key features') te bevatten.⁸ Het is echter de vraag of kennis een goede parameter is voor de kwaliteit van zorg. Men weet niet of de kennis werkelijk in praktijk wordt gebracht en invullen van een kennistoets is wat anders dan het beantwoorden van hulpvragen.

Een vaardigheidstoets, waarin artsen laten zien hoe zij een bepaalde taak uitvoeren, benadert de werkelijkheid beter.⁹ Deze methode richt zich op observeerbaar gedrag en is gestandaardiseerd. Iedere arts krijgt dezelfde opdracht in een identiek gesimuleerde setting, bijvoorbeeld in een skillslab met simulatiepatiënten. De methode is logistiek complex en de voorspellende waarde van simulatietoetsen voor het dagelijks handelen is niet eenduidig. 'Laten zien wat je kunt' is iets anders dan

'dagelijks doen'.¹⁰ In onderzoeken waar een relatie aannemelijk werd gemaakt, was het handelen van de arts beoordeeld door collega's met wie hij samenwerkte, of afgeleid van indirecte indicatoren, zoals verwijs- of voorschrijfcijfers.¹¹⁻¹³

Recent hebben wij een valide observatiemethode ontwikkeld waarmee het dagelijks handelen van huisartsen tijdens consulten direct kan worden getoetst.¹⁴ Deze videotoevoeging maakt het mogelijk nader onderzoek te doen naar de mate waarin andere toetsmethoden, zoals kennistoetsen en een simulatiespreekuur, de werkelijkheid van alledag weerspiegelen. Daarbij is het uiteindelijke doel het ontwikkelen van een integrale toets gericht op kwaliteitsverbetering van praktijkmanagement, arts-patiëntcommunicatie en medisch handelen in de praktijk van alledag, bestaande uit de meest valide en best toepasbare methoden.

De vraagstellingen van dit onderzoek luiden:

- Welke meetkundige verschillen zijn er tussen observatie van het handelen van huisartsen tijdens simulatieconsulten en tijdens echte consulten?
- In hoeverre zijn resultaten behaald op kennis- en simulatietoetsen voorspellend voor de kwaliteit van het dagelijks handelen?

Methode

Integrale toets

Bestaande Nederlandse toetsinstrumenten werden onderzocht op validiteit, betrouwbaarheid en haalbaarheid, en waar nodig verbeterd. In de integrale toets werden opgenomen:

- een Algemeen Medische Kennistoets (AMK, 131 items) en een Kennistoets over Medisch-technische Vaardigheden (KOV, 124 items), beide gebaseerd op de blauwdruk – een kwalitatieve en kwantitatieve inhoudsspecificatie – van de Landelijke Huisartsgeneeskundige Kennistoets;^{15,16}
- een kennistoets over aspecten van arts-patiëntcommunicatie, zoals vraagver-

heldering en informatie geven (KCOM, 12 items); een blauwdruk voor een dergelijke toets ontbrak en het aantal beschikbare vragen was beperkt.

In het kader op deze pagina wordt een voorbeeld gegeven van de drie soorten vragen en de antwoordmogelijkheden.

Het score-instrument en de criterialijst 'MAAS-globaal' werden gebruikt voor de observatie van echte consulten en consulten in een simulatiesprekkuur.^{17,18}

Voor deze integrale toetsing werd een model ontwikkeld waarin vier niveaus worden onderscheiden (figuur):

- kennen;
- weten hoe te doen;
- laten zien;
- doen.

De medische kennistoetsen garanderen een brede 'dekking' van het medische handelen. Aandoeningen uit alle ICPC-hoofdstukken zijn in de toetsen opgenomen: in de AMK op het niveau van 'kennen', en in de KOV op het niveau 'weten hoe te doen'. De KCOM dekt een beperkt aantal aspecten van de arts-patiëntcommunicatie op beide niveaus. Het simulatiesprekkuur meet vaardigheden met betrekking tot arts-patiëntcommunicatie en het medisch handelen op het niveau 'laten zien'. De praktijkvideotoets meet deze aspecten op het niveau dagelijks 'doen'. Bij beide observatiemethoden worden aandoeningen uit minimaal acht ICPC-hoofdstukken getoetst. De plaats van praktijkmanagement in relatie tot deze integrale toets is eerder in H&W beschreven.¹⁹

Deelnemers en onderzoeksdesign

Alle huisartsen in het zuiden van Nederland werden per brief uitgenodigd deel te nemen aan het onderzoek. Een kwart van hen (220) reageerde positief. Om budgettaire redenen werden 135 huisartsen at random geselecteerd; 100 van hen namen deel aan de toetsen en 35 werden getraind in het observeren van consulten. Kenmerken van de huisarts, zoals leeftijd, geslacht, fulltime of parttime werken, wel of niet werkzaam zijn als solist of huisartsop-leider, werden genoteerd om de representativiteit van de onderzoekspopulatie voor

de Nederlandse huisarts en de invloed op de toetsresultaten vast te stellen.²⁰

De 100 huisartsen maakten thuis de drie kennistoetsen. Vervolgens werden zij door loting ingedeeld in twee groepen, die vergelijkbaar waren met betrekking tot de gemeten arts- en praktijkenmerken en de scores op de kennistoetsen. De huisartsen uit groep 1 namen eerst deel aan het simu-

latiesprekkuur op het gedragslab van de Universiteit Maastricht en vervolgens aan de videotoets in de eigen praktijk. Huisartsen uit groep 2 deden hetzelfde, maar in omgekeerde volgorde. Op deze manier was het mogelijk de invloed van de ene toets op het handelen in de tweede toets te analyseren ('volgorde-effect').²¹ De consulten in beide settings werden op video

Drie soorten kennisvragen: algemeen medische vaardigheden, technische vaardigheden en communicatieve vaardigheden

Bewegingsapparaat (ICPC L)

Mevrouw Jorissen komt met haar dochtertje Josien, 4 jaar, op het spreekuur omdat Josien X-benen heeft waardoor ze wat moeilijk loopt en vaak valt. De huisarts constateert bij onderzoek inderdaad genu valgum met een intermalleolaire afstand van 7 cm. De huisarts vertelt dat verwijzing naar een orthopeed op dit moment is geïndiceerd.

1 Dit is correcte informatie.

Literatuur

Streefkerk JG, Verheij ThJM (red). Kleine kwalen in de huisartspraktijk. Utrecht: Bunge, 1997:261-3.

Tractus circulatorius (ICPC K)

Een huisarts ziet een man van ongeveer 60 jaar voor haar op straat neervallen. Ze gaat er naar toe en onderzoekt hem. De man blijkt bewusteloos en er is geen circulatie voelbaar. Zij besluit te reanimeren en hanteert daarbij het volgende schema: na 10 hartmassages telkens 1 beademing.

2 Dit schema is volgens de richtlijnen van de Nederlandse Hartstichting.

Voor wat betreft de frequentie van de hartmassages: ze geeft ongeveer 90 massages in één minuut.

3 Deze frequentie is volgens de richtlijnen van de Nederlandse Hartstichting.

Literatuur

Wanneer elke seconde telt. Nederlandse Hartstichting, 1996:27.

Mevrouw Dellevoort komt bij de huisarts om de laboratoriumuitslag van haar dochtertje Jasmine te bespreken. De huisarts weet dat bij Jasmine leukemie is vastgesteld en moet dit nu aan mevrouw Dellevoort vertellen. Het consult begint als volgt (H=huisarts, M=moeder):

H 1 U komt voor de uitslag van het bloedonderzoek van uw dochtertje?

M 2 Ja, dat klopt.

H 3 Ik heb helaas slecht nieuws voor u: uit het bloedonderzoek blijkt dat Jasmine leukemie heeft.

M 4 Weet u dat wel zeker, is het bloed wel goed onderzocht?

H 5 Ik begrijp dat het voor u een enorme schok is ...

U moet er vanuit gaan dat het bloed goed onderzocht is.

M 6 Maar, moet het dan niet nog een keer herhaald worden?

In de literatuur (bijvoorbeeld Schouten, Anamnese en advies, Samsom Stafleu 1985) worden richtlijnen gegeven ten aanzien van het omgaan met 'slecht nieuws' door de (huis)arts.

4 In regel 3 van de hierboven beschreven casus handelt de huisarts volgens de richtlijnen.

5 De wijze waarop mevrouw Dellevoort op het slechte nieuws reageert (in regel 4 en 6) wordt 'verwarring' genoemd.

opgenomen en door observatoren (n=35) gescoord op arts-patiëntcommunicatie en medisch handelen. De observatoren hadden hiervoor vier trainingen gevolgd.

Simulatiespreekuur (SSU) en praktijkvideotoets (PVT)

Een panel van zeven huisartsen schreef acht casus voor het simulatiespreekuur; de inhoud hiervan werd bepaald door het basistakenpakket, prevalentiecijfers uit de huisartspraktijk en NHG-standaarden waarmee de observatoren het scoren van het medisch handelen konden onderbou-

wen.²²⁻²⁵ Het gedragslab van de Universiteit Maastricht werd omgebouwd tot een groepspraktijk met vier units. In een roulatiesysteem presenteerden getrainde simulatiepatiënten hun problemen aan de huisarts; deze moest het consult binnen twaalf minuten zo natuurlijk mogelijk afhandelen. De onderwerpen van deze casus staan in tabel 1.

Voor de observatie van praktijkconsulten werd bij de huisartsen een video-observatiesysteem geïnstalleerd, bestaande uit twee camera's met ingebouwde microfoon.¹⁴ Gedurende een week nam de huis-

arts in de spreekkamer en in de onderzoeksruimte na toestemming van de patiënt elk consult integraal op en werden consultgegevens genoteerd in een logboek. Om de representativiteit van de consulten en standaardisatie van de steekproeven te bevorderen, werden op basis van prevalenties, het basistakenpakket en NHG-standaarden selectiecriteria geformuleerd. Daarmee werden uit het logboek twintig consulten geselecteerd, zestien voor observatie en vier als reserve.

Evaluatie van toetsmethoden

De huisartsen vulden vragenlijsten in met vragen over de mate van echtheid van het simulatiespreekuur, de invloed van de observaties op hun handelen, de mate van herkenning van hun eigen werkstijl in de verschillende video-opnamen, de acceptatie van de methoden en de mate waarin deze aanzetten tot verbeteractiviteiten.

Analyse

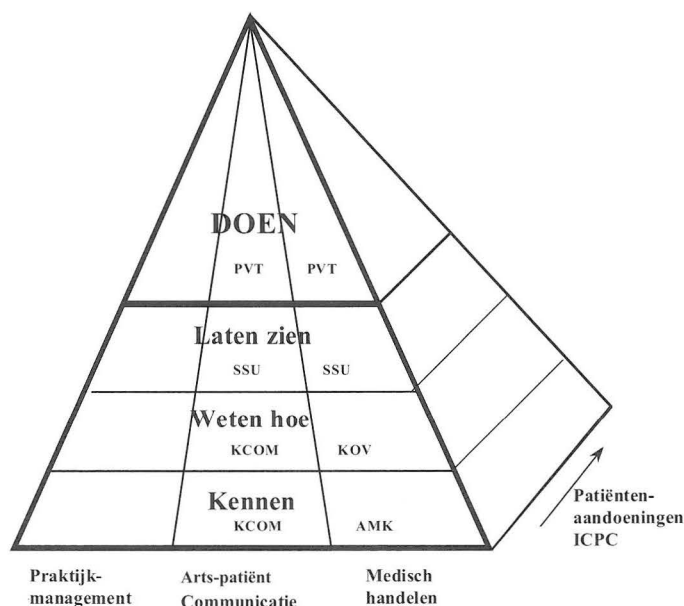
De gemiddelde kennistoetsscores werden berekend over de goed minus fout scores, uitgedrukt in percentages van de maximaal te bereiken score (-100 = alles fout tot +100 = alles goed). Cronbach's alfa (maat voor betrouwbaarheid) werd berekend over de goed-min-fout scores. Met de Spearman-Brown Prophecy Formula werd berekend hoeveel items nodig zijn voor een betrouwbare kennistoets over communicatie (Cronbach's alfa $\geq 0,80$).

De scores voor de consultobservaties werden per huisarts per consult en per steekproef berekend door de itemscores op te tellen en uit te drukken in percentages van het maximaal haalbare voor arts-patiëntcommunicatie, voor medisch handelen en voor beide aspecten samen (som van beide scores gedeeld door 2: de totale score).

De betrouwbaarheid van de observaties werd uitgedrukt in de mate van reproduceerbaarheid, in een coëfficiënt die is gerelateerd aan het aantal casus en het aantal observatoren, rekening houdend met verschillende variantiebronnen, zoals intra- en interobservatorbetrouwbaarheid, de moeilijkheidsgraad van het consult en andere 'ruis' zoals 'je dag niet hebben'.²⁶

Tabel 1 Inhoud van de acht simulatieconsulten (ICPC-hoofdstukken)

1	Controle diabetes mellituse type 2	T
2	Aspecifieke maagklachten	D
3	Inspanningsasthma	R
4	Keelpijn en doofheid	R+H
5	Schouderklachten (painfull arc)	L
6	Slapeloosheid/moeheid	P+B
7	Angina pectoris en claudicatio intermittens	K
8	Hoofdpijn en wazig zien	F



Figuur Blauwdruk integrale toets

De meetkundige kwaliteiten van de observatiemethoden werden met elkaar vergeleken, met name de inhoudsvaliditeit (in hoeverre meet de methode het vakgebied, wat het beoogt te meten), de criteriumvaliditeit (in hoeverre hangen toetsscores samen met de scores van de praktijkvideotoets, het criterium), de betrouwbaarheid en de haalbaarheid (tabel 2).

Resultaten

De onderzoekspopulatie verschilde ten aanzien van de gemeten kenmerken niet van de populatie Nederlandse huisartsen,

met uitzondering van de gemiddelde leeftijd (onderzoekspopulatie was 1,2 jaar jonger, $p < 0,05$). Tien huisartsen vielen uit, met als redenen 'te bedreigend' (1), 'te druk' (5) en 'op vakantie' (4).

De gemiddelde scores behaald in het simulatiespreekuur, de videotoets en de kennistoetsen zijn vermeld in tabel 3. De scores bij de observatiemethoden zijn onderling vergelijkbaar, evenals de kennistoetsscores. Referentiewaarden voor de scores ontbreken, doordat de toetsen in deze vorm nooit eerder zijn afgenomen. De scores kunnen ook niet vertaald worden in rapportcijfers, omdat de items onderling niet zijn gewogen en een grens tus-

sen een voldoende en onvoldoende niet is vastgesteld. De deelnemers konden hun individuele gemiddelde scores op ICPC- of itemniveau vergelijken met de gemiddelde scores van hun groep (relatieve norm).

Meetkundige verschillen

Inhoudsvaliditeit (tabel 2). Groep 2 scoorde hoger in het simulatiespreekuur dan groep 1 ($p < 0,05$). De groepen verschilden niet in de scores voor de praktijkconsulten (tabel 3). Alleen het simulatiespreekuur blijkt dus gevoelig voor een volgorde-effect.

Observatie van praktijkconsulten is in-

Tabel 2 Methoden waarmee psychometrische kenmerken van het simulatiespreekuur en de praktijkvideotoets zijn vergeleken

Kenmerken	Simulatiespreekuur	Praktijkvideotoets
Validiteit		
<i>Inhoudsvaliditeit</i>		
– inhoud van de geselecteerde consulten	casuscriteria: prevalentie/standaarden	selectiecriteria: prevalentie/standaarden
– tijd/volgorde-effect	t-test	t-test
– representativiteit van de consulten	per definitie	statistiek: % steekproeven dat voldoet aan selectiecriteria
– invloed steekproefverschillen op scores	n.v.t. (identieke consulten)	regressieanalyse
– authenticiteit van de consulten	vragenlijst	n.v.t. (impliciet)
– 'audience'-effect	vragenlijst	vragenlijst
– herkenning eigen werkstijl	vragenlijst	vragenlijst
<i>Criterium validiteit</i>	Pearson correlaties	is het criterium
Betrouwbaarheid	generaliseerbaarheidsanalyse	generaliseerbaarheidsanalyse
Haalbaarheid		
– acceptatie	vragenlijst	vragenlijst
– kosten	boekhouden	boekhouden

Toelichting

De *inhoudsvaliditeit* is de mate waarin de inhoud van de toets een weerspiegeling vormt van het domein waarop de toets betrekking heeft. Deze validiteit wordt door de volgende te meten aspecten beïnvloed:

- de inhoud van de consulten (takenpakket en standaarden-gerelateerd, prevalentie);
- het volgorde-effect, verschillen scores tussen beide observatie-methoden binnen en tussen beide groepen (gepaarde en niet-gepaarde t-testen);
- de huisartsgeneeskundige representativiteit, voor simulatieconsulten gegarandeerd door criteria voor script, voor praktijkconsulten afhankelijk van mate van effectuering selectiecriteria (beschrijvende statistiek);
- de invloed van steekproefverschillen op de praktijk-scores (multiple regressie-analyse, scores als afhankelijke en steekproefverschillen als onafhankelijke variabelen);
- de authenticiteit, van de simulatieconsulten

- (vragenlijst voor huisartsen, ja-nee-geen mening) en de praktijkconsulten (van nature echt);
- de invloed van de observaties (audience-effect) op het handelen van huisartsen en de mate van herkenning van de eigen werkstijl (observatie van video-opnames door huisarts en vragenlijst, vijfpunts Likert-schaal).

Criterium-validiteit is de mate waarin scores van toetsmethoden correleren met de scores van een extern criterium, de meest valide toets. In onze studie is dat de observatie van praktijkconsulten. Deze methode meet de feitelijk geboden zorg, het primaire doel van kwaliteitsverbetering. Daaraan werden de scores van het simulatiespreekuur gecorreleerd (ware correlaties = Pearson's correlaties gecorrigeerd voor de mate van onbetrouwbaarheid van de toetsen).

De *betrouwbaarheid* is de mate waarin de toets bij herhaalde afname tot dezelfde resultaten leidt (=

de reproduceerbaarheid). Deze werd uitgedrukt in generaliseerbaarheidscoëfficiënten ($\geq 0,80$ = voldoende betrouwbaar) in relatie tot het aantal consulten en het aantal observatoren (een of twee).

De *toepasbaarheid* werd uitgedrukt in kosten per aantal observaties. De mate waarin de huisartsen de toetsen accepteren en de toetsen aanzetten tot verbeteractiviteiten (= educatieve waarde), werd door beschrijvende statistiek van antwoorden op evaluatievragen in kaart gebracht.

De voorspellende waarde van de kennistoetsen en het simulatiespreekuur voor de kwaliteit van het dagelijks handelen werd berekend met multiple regressie-analyses (scores van praktijkconsulten afhankelijke variabelen; kennistoetsscores, scores voor simulatieconsulten, arts-praktijkkenmerken en volgorde-effect (confounders) als onafhankelijke variabelen).

Tabel 3 Gemiddelde MAAS-Globaal-scores en standaarddeviatie (SD) van simulatie-spreekuur en praktijkvideotoets en de gemiddelde goed-min-fout scores van kennistoetsen

	Groep 1	Groep 2
Simulatiespreekuur communicatie	44,4 (10,3)	52,4 (8,6)
medisch handelen	64,9 (8,5)	70,7 (6,4)
Praktijkvideotoets communicatie	41,3 (7,7)	42,1 (7,1)
medisch handelen	62,8 (7,6)	61,3 (6,2)
Algemene kennis	52,7 (11,5)	50,8 (11,0)
Kennis over vaardigheden	31,1 (10,9)	31,4 (12,5)
Kennis over communicatie	10,8 (33,3)	14,8 (26,4)

Tabel 4 Het percentage huisarsen dat de eigen werkstijl met betrekking tot communicatie en medisch handelen herkent en aangeeft in welke mate men beïnvloed is door de observatie (n=85)

	Groep 1		Groep 2	
	SSU	PVT	SSU	PVT
Herkenning eigen communicatiestijl	6,3	93,7	16,7	83,3
Herkenning werkstijl medisch handelen	6,3	93,7	12,2	87,8
Permanent beïnvloed	75,0	35,5	50,0	22,0
Niet permanent beïnvloed	25,0	64,5	50,0	78,0

SSU=simulatiespreekuur; PVT= praktijkvideotoets.

Tabel 5 Ware correlaties (Pearson) tussen enerzijds praktijkvideotoets en anderzijds kennistoetsen en simulatiesprekkuur

		Groep 1	Groep 2
Medisch	algemene kennis	0,54 (p<0,01)	0,43 (p<0,01)
	kennis over vaardigheden	0,56 (p<0,01)	0,43 (p<0,01)
	simulatiespreekuur	0,59 (p<0,01)	0,33 (p<0,05)
Communicatief	kennis	0,73 (p<0,01)	-0,10 (n.s.)
	simulatiespreekuur	0,17 (n.s.)	0,25 (n.s.)

Tabel 6 Resultaten regressieanalyses met de praktijkvideotoetscores (medisch en communicatief) als afhankelijke variabelen en de scores van kennistoetsen en het simulatiespreekuur en het volgorde-effect als onafhankelijke variabelen

		Bèta	t	R	R ²
Medisch	algemene kennistoets	0,26	2,61*	0,37	0,13
	simulatiespreekuur	0,33	3,10*	0,43	0,19
	volgorde-effect	-0,23	-2,19†	0,48	0,23
Communicatief	kennistoets over communicatie	0,28	2,72*	0,28	0,08

* $p \leq 0,05$;
† $p \leq 0,01$.

houdsvaarde. Alle selectiecriteria zijn in minimaal 70 procent van de steekproeven geëffectueerd. Bovendien hebben verschillen in steekproefkenmerken geen invloed op de scores, met uitzondering van de totale duur van de steekproeven. Naarmate de steekproef langer duurt, zijn de scores hoger. De praktijkconsulten zijn van nature authentiek.¹⁵

Van de huisartsen vond 94 procent de inhoud van de simulatieconsulten huisartsgeneeskundig authentiek, 3 procent vond dat niet en 3 procent oordeelde neutraal. *Tabel 4* laat echter zien dat de invloed van de observatie op de deelnemers in het simulatiespreekuur groter is dan tijdens de observatie in de praktijk. Vergelijken met het simulatiespreekuur herkenden aanzienlijk meer deelnemers hun werkstijl in de video-opnamen van de praktijkconsulten. Deze methode weerspiegelt het dagelijks handelen dus beter dan het simulatiespreekuur.

De scores voor de arts-patiëntcommunicatie (groepen 1 en 2) van het simulatiespreekuur correleren niet met de scores van de praktijkobservatie (tabel 5). De score van de kennistoets over communicatie (KCOM) correleert alleen in groep 1 met de score voor arts-patiëntcommunicatie in de praktijk.

Bij medisch handelen correleren scores van het simulatiespreekuur en de medische-kennistoetsen (in gelijke mate) met de scores voor medisch handelen in de praktijk.

Met andere woorden: het simulatiesprekuur weerspiegelt de dagelijkse arts-patiëntcommunicatie niet en benadert het dagelijks medisch handelen niet beter dan kennistoetsen.

Voorspellende waarde

De voorspellende waarde van het simulatiespreekuur en de kennistoetsen voor het dagelijks handelen (tabel 6) blijkt beperkt. Toetsscores verklaren samen slechts 19 procent van de variantie in het dagelijks medisch handelen en slechts 8 procent van de variantie in dagelijkse arts-patiëntcommunicatie (R^2). Dat betekent dat de voorspellende waarde van deze toetsen voor het dagelijks handelen te gering is om ob-

servatie van praktijkconsulten te kunnen vervangen.

De bijdrage van het simulatiesprekkuur aan de verklaarde variantie is gering: na de algemeen medisch kennistoets voegt het simulatiesprekkuur nog 6 procent toe aan de verklaarde variantie (R^2 neemt toe met 0,6).

De kennistoets over communicatie (12 items) draagt in bescheiden mate bij aan de verklaarde variantie in arts-patiëntcommunicatie in de praktijk; het simulatiesprekkuur draagt niets bij.

De voorspellende waarde van een gestandaardiseerd simulatiesprekkuur voor het medisch handelen in de praktijk is minder dan die van kennistoetsen; voor communicatie is deze nihil.

Artskenmerken zoals leeftijd, geslacht en praktijkvorm dragen niet bij aan de verklaring van variantie in scores voor het dagelijks handelen.

Betrouwbaarheid

De Cronbach's alfa van de algemeen medische kennistoets is 0,64, van de kennistoets over medisch-technische vaardigheden 0,60, en die van de kennistoets over communicatie (KCOM) 0,45. Een KCOM blijkt 50 items nodig te hebben om voldoende betrouwbaar te zijn.

De totale scores (= scores voor communicatie + medisch handelen) van beide observatiemethoden zijn betrouwbaar (generaliseerbaar) als acht consulten door twee observatoren (= 16 observaties) of twaalf consulten door één observator (= 12 observaties, totaal 3 uur) zijn gescoord. Voor arts-patiëntcommunicatie in simulatieconsulten zijn deze aantallen hetzelfde; voor echte praktijkconsulten zijn dat er meer: twaalf consulten door twee observatoren of zestien door één observator (= 4 uur observeren).

Voor medisch handelen zijn deze aantallen in het simulatiesprekkuur: twintig consulten bij één observator en zestien bij twee observatoren; in de praktijk: zestien consulten met één observator en twaalf met twee, dat wil zeggen vier observaties (= 1 uur) minder dan bij het simulatiesprekkuur.

Toetsing van echte consulten uit de

praktijk is dus voor communicatie wat minder en voor medisch handelen wat meer betrouwbaar dan toetsing in een gestandaardiseerd simulatiesprekkuur.

Haalbaarheid

De observatiemethoden hebben een aantal gelijke kostenposten, met name administratie, porto, aanschaf van videobanden, observatortraining, honorarium voor observatoren, data-invoer, data-analyse en feedback.

Beide methoden hebben daarnaast specifieke kosten:

Voor het simulatiesprekkuur zijn dat huur van het skillslab, honoraria voor technicus, simulatiepatiënten en auteurs van de simulatiecasus, training van simulatiepatiënten, reiskosten voor huisartsen en simulatiepatiënten, en waarneemkosten (huisartsen zijn uit de praktijk).

Specifieke kosten voor de videotoets zijn: aanschaf video-apparatuur (een set per tien huisartsen), training, honorarium en reiskosten video-installateurs, en selectie van consulten uit het logboek.

De kosten voor de kennistoetsen zijn beperkt: honorarium voor auteurs en corrector, drukkosten, data-invoer, data-analyse en feedback.

In dit onderzoek kostte het simulatiesprekkuur NLG 1050 en de videotoets NLG 750 per huisarts (observatie van 12 consulten). Indien meer consulten worden geobserveerd, stijgen de kosten voor het simulatiesprekkuur drie keer zo snel als de kosten van de videotoets, die grotendeels stabiel zijn (aanschaf en installatie video-apparatuur versus instabiele kosten voor het simulatiesprekkuur, zoals huur, honoraria voor de technicus en de simulatiepatiënten en waarneemkosten). Bij observatie van zestien consulten kost het simulatiesprekkuur NLG 1500 en de videotoets NLG 900 per deelnemer. Kennistoetsen kostten ongeveer NLG 50 per kennistoets met 150 items (een item-bank was voorhanden).

Acceptatie en educatieve waarde

De feedback per toetsmethode was beperkt om de voorkomen dat scores van opeenvolgende toetsen onderling beïnvloed

zouden worden (preventie volgorde-effecten). Dit vergrootte het risico op een verminderde acceptatie en op een lagere educatieve prikkel (impuls tot verbeteractiviteiten).

De huisartsen is gevraagd zelf een integrale toets samen te stellen voor arts-patiëntcommunicatie en medisch handelen (vragenlijst, respons 78%). Voor medisch handelen zijn de kennistoetsen en de videotoets favoriet, voor communicatie is de videotoets het meest gewaardeerd.

Verdere evaluatie (ja/nee-vragen en vijf-punts Likert-schaal: 1 = geheel oneens; 5 = geheel eens) gaf het volgende beeld:

- Gelet op de inspanning van deelname heeft men meer profijt van observatiemethoden dan van kennistoetsen: 3,9 versus 3,5.
- Kennistoetsen geven bij iets meer huisartsen inzicht in sterke en zwakke kanten (kennis 3,7, simulatiesprekkuur 3,5, videotoets 3,4).
- Observatiemethoden geven meer inzicht in arts-patiëntcommunicatie: videotoets 4,0, simulatiesprekkuur 3,9, kennis 2,9.
- Deelname aan de kennistoets leidt bij meer huisartsen tot concrete veranderingen in het dagelijks medisch handelen dan de observatiemethoden doen: 41 versus 22 en 24 procent (respectievelijk simulatiesprekkuur en videotoets). Voor arts-patiëntcommunicatie zijn deze percentages 15, 39 en 55 procent.
- Ten slotte heeft deelname en feedback bij een beperkt aantal huisartsen geleid tot concrete nascholingsactiviteiten, voor medische onderwerpen 35, 17 en 17 procent, en voor arts-patiëntcommunicatie 9, 19 en 25 procent (respectievelijk kennistoets, simulatiesprekkuur en videotoets).

De huisartsen hebben een voorkeur voor medische kennistoetsen en voor observatie van consulten in de praktijk. De educatieve prikkel tot concrete veranderingen in de dagelijkse praktijk is bij medische kennistoetsen en videotoets redelijk; de deelname aan reguliere deskundigheidsbevordering wordt zeer beperkt gestimuleerd.

Beschouwing

In dit onderzoek is de voorspellende waarde van een simulatietoets en van kennis-toetsen voor de kwaliteit van het dagelijks handelen vastgesteld, en zijn de meetkundige eigenschappen van twee observatiemethoden systematisch met elkaar vergeleken. Kennistoetsen blijken het dagelijks handelen in dezelfde mate en soms beter te voorspellen dan een gestandaardiseerd simulatiespreekuur, dat in meerdere opzichten de mindere is van observatie van het dagelijks handelen in de praktijk.

De huisartsen namen vrijwillig deel aan het onderzoek, hetgeen de acceptatie en de educatieve waarde van de verschillende toetsmethoden kan hebben beïnvloed. De methoden zijn echter binnen dezelfde populatie met elkaar vergeleken, hetgeen de geldigheid van de conclusies ondersteunt.

De verschillen tussen de meetkundige eigenschappen van het simulatiespreekuur en de videotoets zijn markant. De gevoeligheid van het simulatiespreekuur voor een volgorde-effect, dat leidt tot hogere scores, en het ontbreken van een samenhang tussen beide methoden betreffende de arts-patiëntcommunicatie zijn te herleiden tot dezelfde noemer: het simulatiespreekuur is een competentietesting, het meet wat huisartsen kunnen en niet wat zij gewend zijn te doen. De hoge mate waarin huisartsen zich beïnvloed voelen in deze setting en de beperkte mate waarin zij daarin hun gebruikelijke werkstijl herkennen, versterken deze conclusie. Het onnatuurlijke karakter van de simulatiesetting nodigt uit tot ander gedrag.

Voor de arts-patiëntcommunicatie wordt daardoor beïnvloed. Op het moment dat de huisarts zich vakinhoudelijk richt op anamnese en lichamelijk onderzoek, vergeet hij de bijzondere omgeving en handelt hij meer natuurgetrouw. Deze effecten kunnen per huisarts en per simulatiecasus verschillen, waardoor huisartsen minder consistent handelen dan in de eigen praktijk. Dit heeft negatieve consequenties voor de betrouwbaarheid van deze methode, die op deze wijze minder gestandaardiseerd is dan men zou wensen.

In de praktijk van alledag, in de eigen

omgeving met bekende patiënten, handelen huisartsen consistent. De huisartsen herkennen hun werkstijl dan ook vooral in video-opnamen van echte consulten. Dit consistente gedrag en de standaardisatie door de selectiecriteria¹⁴ bevorderen de inhoudsvaliditeit en de generaliseerbaarheid van deze methode: na drie uur observatie kan een betrouwbare uitspraak worden gedaan over de kwaliteit van het medisch en communicatief handelen van de huisarts. De toetslengte van betrouwbare simulatiemethoden is vaak aanmerkelijk langer.

De lage voorspellende waarde voor het dagelijks handelen, de hoge kosten en de beperkte mate van acceptatie door de deelnemers maken het simulatiespreekuur niet geschikt voor gebruik op grote schaal. Praktijkbezoek door simulatiepatiënten of simulatietoetsen in een skillslab zijn nuttig voor training of als effectmeting van (na)scholing, met het voordeel dat de simulatiepatiënten de arts feedback kunnen geven.

Kennistoetsen zijn minder bewerkelijk en voorspellen het dagelijks medisch handelen even goed als het simulatiespreekuur. Voor arts-patiëntcommunicatie is deze voorspellende waarde zelfs beter.

De toetsen werden thuis gemaakt, waarbij de deelnemers werd gewezen op de educatieve waarde als er niet werd gespiekt: direct invullen weerspiegelt (lacunes in) parate kennis. Literatuurverwijzingen werden pas achteraf toegezonden. Bijna 90 procent van de huisartsen maakte de toetsen binnen 3,5 uur,²⁷ wat het onwaarschijnlijk maakt dat op grote schaal antwoorden zijn opgezocht (bij bijna driehonderd items). Ook de samenhang tussen de kennistoetsscores en de scores voor de praktijkconsulten, waar parate kennis moet worden toegepast, duidt op niet 'spieken'.

De huisartsen accepteerden de kennistoetsen goed, wat een reden te meer is om de betrouwbaarheid van deze toetsen te verbeteren. Daarbij valt bijvoorbeeld te denken aan het toetsen van thema's en het stellen van een absolute norm (voldoende versus onvoldoende), waardoor enerzijds

huisartsen en anderzijds onderwerpen kunnen worden geselecteerd voor kwaliteitsverbetering.

Een kennistoets over arts-patiëntcommunicatie moet nog worden ontwikkeld. Tevens dienen absolute normen gesteld te worden voor communicatief en medisch handelen. Het vergelijken van eigen scores met die van de groep (relatieve norm) prikkelt huisartsen, maar beantwoordt niet de vraag of de kwaliteit voldoende is. Pas dan kunnen de lagere scores voor communicatie in ons onderzoek werkelijk op hun waarde geschat worden.

Ook moet worden onderzocht op welke wijze de huisarts ondersteund kan worden in het vertalen van feedback in effectieve verbeterplannen en uitvoering daarvan. Toetsing zonder een zorgvuldig natraject is verspilde energie.

Conclusie

Integrale toetsing van arts-patiëntcommunicatie en medisch handelen impliceert observatie van echte consulten in de praktijk en, gezien eerdere bevindingen, toetsing van praktijkmanagement door visitatie. Gelet op de relatief lage kosten en de voorspellende waarde voor het feitelijk handelen zijn kennistoetsen second best.

Literatuur

- 1 Newble D, Jolly B, Wakeford R. The certification and recertification of doctors: issues in assessment of clinical competence. Cambridge: Cambridge University Press, 1994.
- 2 Kwaliteit op koers. Utrecht: Landelijke Huisartsen Vereniging en Nederlands Huisartsen Genootschap, 1999.
- 3 Berwick DM. Continuous improvement as an ideal in health care. *New Engl J Med* 1989;320: 53-6.
- 4 Berwick D. Heal thyself or heal the system: can doctors help to improve medical care? *Quality Health Care* 1992;1(Suppl):S2-S8.
- 5 Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;1:41-67.
- 6 Kane MT. The assessment of professional competence. *Evaluation & the health professions* 1992;15:163-82.
- 7 Reznick RK, Smee S, Baumber JS, et al. Guidelines for estimating real cost of an objective structured clinical examination. *Acad Med* 1993;68:

- 513-17.
- 8 Van Leeuwen YD. Growth in knowledge of trainees in general practice. Figures on facts [Dissertatie]. Maastricht: Universiteit Maastricht, 1995.
- 9 Rethans JJ, Westin S, Hays R. Methods for quality assessment in general practice. *Fam Pract* 1996;5:468-76.
- 10 Rethans JJ. Does competence predict performance? Standardized patients as a means to investigate the relationship between competence and performance of general practitioners. Amsterdam: Thesis, 1991.
- 11 Ramsey PG, Carline JD, Inui YS, et al. Predictive validity of certification by the American Board of Internal Medicine. *Ann Intern Med* 1989;110:719-26.
- 12 Norman GR, Davis DA, Painvin A, et al. Comprehensive assessment of clinical competence of family-general physicians using multiple measures. Proceedings 28th conference on research in medical education. Washington: American Association of Medical Colleges, 1989.
- 13 Norman G. Can an examination predict competence? The role of recertification in maintenance of competence. *Annals RCPSC* 1991;24:121-4.
- 14 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;42:439-46.
- 15 Jansen JJM, Tan LHC, Van der Vleuten CPM, et al. Assessment of competence in technical clinical skills of general practitioners. *Med Educ* 1995;29:247-53.
- 16 Van Leeuwen YD, Pollemans MC, Mol SS, et al. The Dutch knowledge test for general practice: issues of validity. *Eur J Gen Pract* 1995;1:113-7.
- 17 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.
- 18 Van Thiel J, Van Dalen J, Ram P. MAAS-globaal criterialijst [Interne publicatie]. Maastricht: Universiteit Maastricht, 1995.
- 19 Ram P, Van den Hombergh P, Grol R, et al. Structuur en proces. Een onderzoek naar de samenhang tussen praktijkvoering en medisch en communicatief handelen door huisartsen. *Huisarts Wet* 1999;42:493-8.
- 20 Caulford PG, Lamb SB, Kaigas TB, et al. Physician incompetence: specific problems and predictors. *Acad Med* 1994;69(Suppl 10):16-20.
- 21 Neufeld VR, Norman GR, redactie. Assessing clinical competence. New York: Springer, 1985.
- 22 Springer MP, redactie. Basistakenpakket van de huisarts. Utrecht: Landelijke Huisartsen Vereniging, 1983.
- 23 Rutten GE, Thomas S, redactie. NHG-Standaarden voor de huisarts. Delen I en II. Utrecht: Bunge, 1993 en 1996.
- 24 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.
- 25 Van de Lisdonk EH, Van den Bosch WJHM, Huygen FJA, Lagro-Jansen ALM. Ziekten in de huisartspraktijk. Utrecht: Bunge, 1994.
- 26 Brennan RL. Elements of generalizability theory. Iowa City: American College Testing Publications, 1983.
- 27 Ram P. Eindverslag project integrale toetsing [Interne publicatie]. Maastricht: Universiteit Maastricht, 1997.