

Diagnostisch toetsoverleg (DTO) vermindert overbodig gebruik aanvullende diagnostiek door huisartsen

WH Verstappen, T van der Weijden, J Sijbrandij, I Smelee, J Hermesen, J Grimshaw, RPTM Grol

Samenvatting

Verstappen WH, Van der Weijden T, Sijbrandij J, Smelee I, Hermesen J, Grimshaw J, Grol RPTM. Diagnostisch toetsoverleg (DTO) vermindert overbodig gebruik aanvullende diagnostiek door huisartsen. *Huisarts Wet* 2004;47(3):127-32.

Achtergrond Huisartsen vragen steeds meer diagnostische tests aan, terwijl volgens de NHG-Standaarden die aantallen juist zouden moeten dalen.

Doel Evaluatie van een innovatieve strategie, het Diagnostisch toetsoverleg (DTO), om het aanvraagdgedrag van huisartsen te veranderen. Wij onderzochten welk effect het DTO heeft op het aantal tests en een aantal volgens de standaarden 'irrationele' tests.

Method Een multicentertrial met een gebalanceerd, incompleet blockontwerp met randomisatie op HAGRO-niveau. Zesentwintig HAGRO's met in totaal 174 huisartsen uit 5 regio's met een diagnostisch centrum vormden de onderzoekspopulatie. Dertien HAGRO's kregen DTO voor cardiovasculaire problemen, bovenbuikklachten en onderbuikklachten, en de andere 13 HAGRO's voor COPD/astma, vage klachten/moeheid en degeneratieve gewrichtsafwijkingen.

Interventie In een halfjaar ontvingen de huisartsen drie opeenvolgende feedbackrapporten waarin hun eigen aanvraagdgedrag in vergelijking met dat van hun collega's in de vorm van een grafiek werd weergegeven. Tijdens drie intercollegiale toetsingsbijeenkomsten in hun HAGRO relateerden zij die rapporten aan bestaande richtlijnen en bespraken de problemen bij en de mogelijkheden tot verandering. Met het maken van expliciete (individuele en HAGRO-) veranderplannen werden de bijeenkomsten afgesloten.

Uitkomstmaten Het gemiddelde van het totale aantal tests, het aantal tests per klinisch probleem en het aantal volgens de richtlijnen 'irrationele' tests.

Resultaten De huisartsen die het DTO kregen voor de klinische problemen A1-3, vroegen per halfjaar gemiddeld 67 tests minder aan dan de controlehuisartsen (95%-BI -104 - -30). Deze huisartsen vroegen ook significant minder 'irrationele' tests voor de klinische problemen A1-3 aan (95%-BI -27 - -7). De huisartsen die de interventie over de klinische problemen B1-3 kregen, vroegen 28 tests minder aan dan de artsen in de controlearm (95%-BI -74 -14).

Conclusie Deze meervoudige en innovatieve interventie maakte gebruik van sociale beïnvloeding onder collega's en bleek een

effectieve methode om het aanvraagdgedrag van huisartsen meer in overeenstemming met de richtlijnen te brengen.

Universiteit Maastricht, Centre for Quality of Care Research (WOK), Care and Public Health Research Institute (Caphri), Capaciteitsgroep Huisartsgeneeskunde: W.H.J.M. Verstappen, huisarts, medisch coordinator; dr. T. van der Weijden, epidemioloog; prof.dr. R.P.T.M. Grol, hoogleraar Kwaliteit van Zorg; Capaciteitsgroep Methodologie en Statistiek: J. Sijbrandij, statistisch medewerker; Elkerliek Ziekenhuis, Stichting voor Diagnostiek en Consultatie, Helmond; dr. I. Smelee, huisarts; Canisius-Wilhelmina Ziekenhuis, Medisch Diagnostisch Centrum, Nijmegen; J. Hermesen, medisch coordinator; University of Ottawa, Ottawa Health Research Institute, Centre of Best Practice, Institute of Population Health, Canada; J. Grimshaw, professor of Public Health.

Correspondentie: wim.verstappen@hag.unimaas.nl

Dit artikel is een aangepaste versie van 'Effect of a practice-based strategy on test ordering performance of primary care physicians: a randomized trial' (JAMA 2003;289:2407-12).

Mogelijke belangenverstrengeling: niets aangegeven.

Inleiding

Hoewel huisartsen volgens geaccepteerde evidence-based richtlijnen te veel diagnostische tests aanvragen, neemt het aantal aangevraagde tests in vele westerse landen, inclusief Nederland, nog steeds toe.¹⁻³ Voor de 250 huisartsen die gebruikmaken van de diensten van het Diagnostisch Centrum Eindhoven bedraagt de stijging van het aantal aanvragen ongeveer 3,3% per jaar vanaf 1999. Een verklaring zou kunnen zijn dat huisartsen een ingeslepen – en dus moeilijk te veranderen – gewoonte hebben bepaalde tests aan te vragen of dat zij bang zijn belangrijke diagnoses te missen of fouten te maken. Het zou ook kunnen liggen aan een gebrek aan kennis over het juist gebruik van diagnostische tests.⁴⁻⁷ De vergrijzing van de bevolking en de verplaatsing van zorg van de tweede naar de eerste lijn lijken voor een verdere toename te zorgen. Ook vragen patiënten vaker dan vroeger expliciet om diagnostische tests en hechten zij vaak meer waarde aan het nut van deze onderzoeken dan op zijn plaats lijkt.^{8,9}

Naar effectieve methoden om het aanvraagdgedrag van huisartsen te veranderen is ook in Nederland al veel onderzoek gedaan. Zo zijn de effecten van een beperkt en een probleemgeoriënteerd laboratoriumformulier onderzocht.^{10,11} Ook is gebruikgemaakt van beslissingsondersteunende systemen en persoonlijke feedback

door een gerespecteerde deskundige.¹²⁻¹³ Omdat uit literatuurreviews blijkt dat sociale beïnvloeding wel eens een veelbelovende strategie zou kunnen zijn, werd een meervoudige, gecombineerde strategie ontworpen: het Diagnostisch toetsoverleg (DTO), dat enigszins voortborduurt op het FTO.¹⁴⁻¹⁶ Huisartsen konden daarin horen en bediscussiëren hoe collega's omgingen met problemen bij het aanvragen van diagnostiek.^{17,18} In de regio Eindhoven was al ervaring opgedaan met intercollegiale toetsgroepen met betrekking tot aanvullende diagnostiek.¹⁹ Het doel van de interventie was een blijvende verbetering van het aanvraaggedrag te bewerkstelligen. We beschrijven hier welke veranderingen het DTO teweegbrengt in de aantallen tests die huisartsen aanvragen.

Methode

Setting en populatie

Vanaf 1 januari 1999 tot 1 juli 2001 werd het onderzoek uitgevoerd in vijf regio's (Kempen, Helmond, Sittard, Nijmegen, de Gelderse Vallei) met een diagnostisch centrum (DC). Een DC is een instelling waar huisartsen diagnostische tests (laboratorium-, röntgen- en functietests) kunnen aanvragen zonder de patiënt naar de tweede lijn te verwijzen. De deelnemende DC's gebruikten voor het aanvragen van laboratoriumonderzoek vergelijkbare probleemgeoriënteerde aanvraagformulieren. Een belangrijke taak van de medisch coördinator van het DC is het geven van feedback aan de betrokken huisartsen over hun aanvraaggedrag. De medisch coördinatoren nodigden de 37 HAGRO's die samenwerkten met deze vijf DC's uit om deel te nemen aan het onderzoek.

Interventie

Literatuuronderzoek en bijeenkomsten met experts resulteerden in de interventiemethode. Het DTO bestond uit de volgende elementen:

- ▶ Er werd persoonlijke feedback gegeven over het eigen aanvraaggedrag in vergelijking met dat van collega's uit de HAGRO en de regio (in de vorm van een grafiek);
- ▶ Bestaande evidence-based richtlijnen werden verspreid over het aanvragen van tests bij bepaalde klinische problemen;
- ▶ Intercollegiale toetsing werd geïntroduceerd in bestaande HAGRO's.

Tabel 1 geeft de klinische problemen met bijbehorende diagnostische tests waar het DTO zich op richtte. Deze tests maken ongeveer 90% uit van de laboratorium-, röntgen- en functietests die huisartsen in een DC kunnen aanvragen.

In de eerste helft van 1999 ontvingen de deelnemende huisartsen driemaal een feedbackrapport over de drie klinische problemen in hun onderzoeksgroep. Tegelijkertijd ontvingen zij daarbij ook een samenvatting van de richtlijnen voor dat specifieke klinische beeld uit NHG-Standaarden en het *Diagnostisch Kompas*. In de onderzoeksperiode kwam de HAGRO onder voorzitterschap van de medisch coördinator van het DC driemaal bij elkaar. De anderhalf uur durende, geaccrediteerde en gestructureerde intercollegiale toetsingsbijeenkomst vond twee weken na de toezending

Wat is bekend?

- ▶ Volgens bestaande evidence-based richtlijnen vragen huisartsen te veel diagnostische tests aan.
- ▶ Er is al veel onderzoek gedaan om de kwaliteit van het huisartsgeneeskundig diagnostisch handelen te verbeteren: de waarde van intercollegiale toetsing is nog onvoldoende onderzocht.

Wat is nieuw?

- ▶ Het DTO (Diagnostisch toetsoverleg) blijkt een effectieve manier om het aantal overbodige diagnostische tests van huisartsen te verminderen.
- ▶ Het DTO is goed te implementeren in de dagelijkse praktijk.
- ▶ Sociale beïnvloeding door collega's heeft een positief effect op de kwaliteit van het diagnostisch handelen van huisartsen.
- ▶ Toepassing van de DTO-methode op andere domeinen van het huisartsgeneeskundig handelen zoals prescriptiecijfers of verwijs cijfers verdient verder onderzoek.

van het rapport en de richtlijnen plaats. Tijdens de bijeenkomsten bediscussieerden en vergeleken huisartsen hun aanvraaggedrag met elkaar en legden zij relaties met de richtlijnen. Ook besluitkundige principes kwamen aan bod zoals de waarschijnlijkheid van fout-positieve uitslagen in een huisartsenpopulatie met lage prevalentiecijfers. Zeker zo belangrijk was de discussie over hoe om te gaan met 'onzinnige' verzoeken van patiënten om een bepaalde test te ondergaan. Sommige huisartsen hadden hier effectieve methodes voor ontwikkeld en deze bleken vaak prima toepasbaar. In het tweede deel van de bijeenkomst kwamen de weerstanden tegen verandering expliciet aan bod en werd geprobeerd daar oplossingen voor te vinden door met collega's te discussiëren waarom en hoe het sommigen desondanks gelukt was om deze weerstanden te overwinnen. Aan het eind was er tijd ingeruimd om te komen tot plannen van individuele huisartsen en HAGRO's om het aanvraaggedrag te veranderen. In de volgende bijeenkomsten werden die plannen geëvalueerd.

Opzet en metingen

Het DTO werd de eerste helft van 1999 geëvalueerd in een multicentertrial met een gebalanceerd, incompleet blockontwerp. De HAGRO's werden gestratificeerd voor groepsgrootte en regio en vervolgens gerandomiseerd (*figuur*). De HAGRO's uit arm I kregen DTO over tests behorend bij klinische problemen, namelijk cardiovasculaire klachten en aandoeningen, bovenbuikklachten en onderbuikklachten (A1-3), terwijl de HAGRO's uit arm II DTO kregen over tests behorend bij de klinische problemen COPD/astma, vage klachten/moeheid en degeneratieve gewrichtsafwijkingen (B1-3) (*tabel 1*). De HAGRO's uit arm I functioneerden als geblindeerde controlegroep voor de HAGRO's uit arm II die de DTO-interventie op de klinische problemen B1-3 ondergingen, en omgekeerd waren de HAGRO's uit arm II de geblindeerde controlegroep voor de interventie op de klinische problemen A1-3.

Dit ontwerp heet *gebalanceerd* omdat de interventie in de twee

Tabel 1 Rationele en irrationele tests bij klinische problemen

probleem	Arm I rationele test	irrationele test	probleem	Arm II rationele test	irrationele test
A1 Cardiovasculaire klachten en aandoeningen	cholesterol, HDL, TG, Na, K, creatinine, (inspannings-)ECG	ureum	B1 COPD/astma	Phadiatop, X-thorax	totaal IgE
A2 Bovenbuikklachten	ALAT, γ -GT, ECHO- bovenbuik	ASAT, LDH, amylase, bilirubine, alkalische fosfatase	B2 Vage klachten/moeheid	BSE, Hb, Ht, MCV, MCHC, MCH, TSH, monospot	aantal leukocyten
A3 Onderbuikklachten	PSA, CRP, coloninloop, ECHO-onderbuik, IVP, sigmoidoscopie		B3 Degeneratieve gewrichtsafwijkingen	BSE, urinezuur reumafactoren	X-lwk, X-cwk, X-schouder, X-knie, X-heup

armen gelijk was en *incompleet* omdat niet alle deelnemende HAGRO's DTO over alle klinische problemen kregen. Een effect op de geïntervenieerde tests van de interventiearm ten opzichte van de controlearm geeft aan dat de DTO-interventie werkelijk effect had en niet het gevolg was van de aandacht die de huisartsen kregen.

Het ontwerp neutraliseert de invloed van enkele niet-specifieke effecten. Zo werd de invloed van het Hawthorne-effect – het effect dat huisartsen hun gedrag veranderen omdat ze in een onderzoek participeren – op deze manier gelijk over beide armen verdeeld. De artsen gaven informed consent om al hun aanvraaggegevens anoniem te gebruiken. Om seizoensinvloeden in het aanvraaggedrag te vermijden werden de voor- en nameting over dezelfde periode van het jaar gedaan: de tweede helft van 1998 en 1999.

Effectmaten

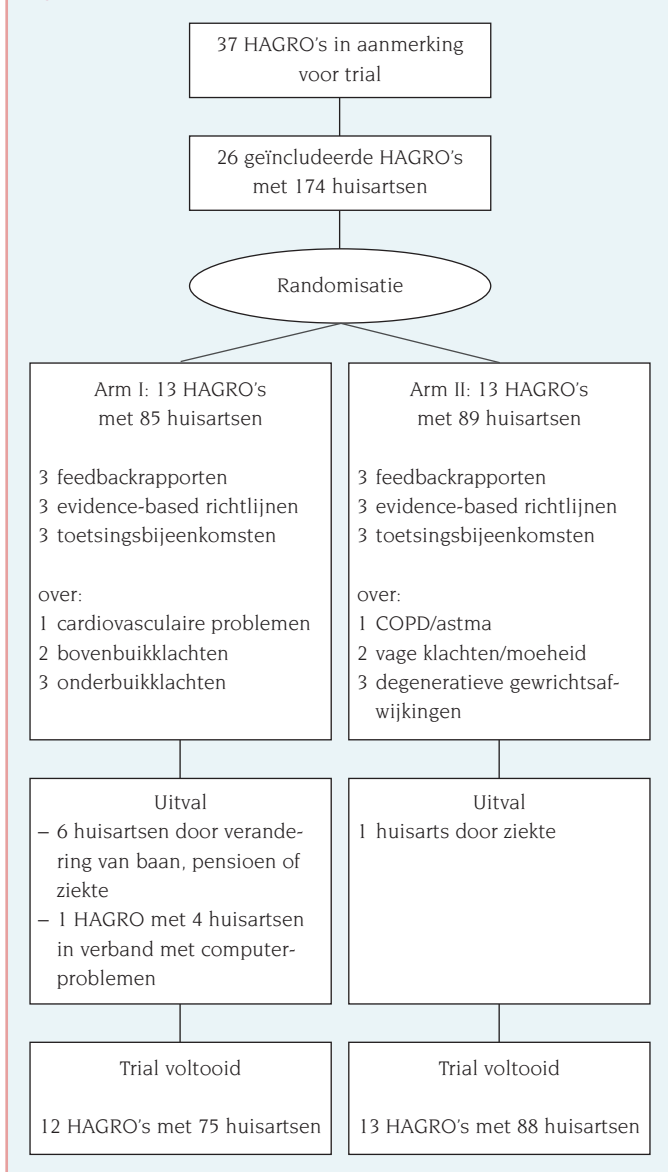
De eerste effectmaat bestond uit het totaal aantal tests per huisarts per halfjaar, en het aantal tests per klinisch probleem per huisarts per halfjaar. De NHG-Standaarden geven in het algemeen aan dat voor de ingesloten klinische problemen te veel tests worden aangevraagd. Een daling in het totale aantal tests beschouwden we daarom als kwaliteitsverbetering. Deze effectmaat analyseerden we ook per klinisch probleem.

Het aantal tests dat de richtlijnen als 'irrationeel' betitelen, was de tweede effectmaat (*tabellen 1 en 2*). Deze tests worden als irrationeel beschouwd bijvoorbeeld omdat het resultaat van de test nauwelijks of geen invloed heeft op de behandeling, omdat er betere alternatieven zijn of omdat de test te veel nadelige gevolgen heeft.

Statistische analyse

Om de clustering binnen HAGRO's mee te kunnen wegen, werd een model met drie niveaus gebruikt met de HAGRO als niveau 3, individuele huisartsen als niveau 2 en de aantallen tests als niveau 1. De analyses werden gedaan met SAS PROC MIXED, versie 8.2 (SAS Institute, Cary, NC). Alle effecten werden geanalyseerd met covariantieanalyse (ANCOVA) met de aantallen tests uit de nameting als afhankelijke variabelen en als onafhankelijke variabelen de aantallen tests uit de nulmeting en de regio omdat de regio een belangrijke voorspeller van het aanvraaggedrag bleek te zijn.²⁰ Deze regressievergelijking geeft het effect β : de totale verandering tussen nulmeting en follow-up van het gemiddeld aantal tests in de interventiearm minus de totale verandering tussen nulmeting en follow-up van het gemiddeld aantal tests in de controlearm, gecorrigeerd voor de aantallen in de nulmeting en de regio. De intracusterrelatiecoëfficiënt (ICC) is een maat om de samenhang van gegevens binnen bepaalde clusters, bijvoorbeeld HAGRO's, weer te geven en geeft de variantie binnen een cluster

Figuur Onderzoeksopzet



Tabel 2 Voorbeelden van delen van richtlijnen over met name enkele 'irrationele' tests

<i>Bovenbuikklachten</i>	Bij vage bovenbuikklachten en algehele malaise (zonder icterus) bestaat er geen reden tot het verrichten van bloedonderzoek op leverfuncties. Het risico van uitgebreide (test)panels is de toenevende kans op fout-positieve waarden. De prevalentie van leverlijden is laag: 4-5 per 1000. Zonder ernstig vermoeden van leverpathologie is er bij een dergelijke lage prevalentie een grote kans op het vinden van fout-positieve uitslagen. Is screenen desondanks nodig dan: - bij moeheid geen leverfuncties; - bij patiënten zonder icterus alleen ALAT en γ-GT; - bij een icterische patiënt een totaal bilirubine, ALAT en γ-GT.
<i>Vage klachten/moeheid</i>	Doe bij aanhoudende vage klachten pas een maand na het eerste consult bloedonderzoek en bepaal dan de BSE, Hb, glucose en TSH. De laatste bepaling vooral bij oudere patiënten. Door 4 weken te wachten wordt de voorafkans op een aandoening aanmerkelijk hoger. De bepaling van het aantal leukocyten geeft nauwelijks meer informatie dan het gegeven wel of geen koorts en geeft geen zekerheid over bacteriële of virale oorzaak. Leukocytdifferentiatie is alleen van belang bij vermoeden van ziekte van Pfeiffer en bij diagnostiek hematologische maligniteiten.
<i>Degeneratieve afwijkingen bewegingsapparaat</i>	Consensus standaardwerkgroepen: voor schouder, heup en knie geldt dat beleid bepaald wordt door het klinisch probleem en niet door de uitslag van de röntgenfoto. Betekenis van verkalkingen is onduidelijk.

Daling in aantallen tests

Tabel 4 laat zien dat de daling van het aantal aangevraagde tests in de interventiearm steeds groter was dan in de controlearm. De huisartsen in arm I vroegen na de interventie 12% minder tests aan dan tijdens de nulmeting, terwijl in arm II geen verandering optrad. In arm I nam het aantal aanvragen met gemiddeld 67 tests meer af per huisarts per halfjaar dan in arm II (95%-BI: -104 - -30). Voor de klinische problemen B1-3 vroegen de huisartsen in arm II 8% minder tests aan ten opzichte van de nulmeting, terwijl de huisartsen in arm I, die hier als controle functioneerden, 3% minder tests aanvroegen. De daling in arm II kwam overeen met een extra daling van gemiddeld 28 tests per huisarts per halfjaar vergeleken met de huisartsen uit arm I (95%-BI: -74 - 14). Tabel 4 laat ook de resultaten per klinisch probleem zien. De gemiddelde daling van het aantal aangevraagde tests uit arm I was statistisch significant voor de drie klinische problemen A1-3, terwijl de daling van het aantal tests per klinisch probleem B1-3 weliswaar overeenkwam met de aanbevelingen uit de richtlijnen, maar zonder statistische significantie.

Daling in aantallen 'irrationele' tests

Tabel 5 laat zien dat huisartsen uit arm I na de interventie significant minder 'irrationele' tests voor de klinische problemen A1-3 aanvroegen (95%-BI -27 - -7). Huisartsen uit arm II vroegen volgens de richtlijnen minder 'irrationele' tests aan voor de klinische problemen B1-3, maar dit verschil was niet statistisch significant (95%-BI: -18 - 2). Er was een significante daling van 4 'irrationele' tests uit het klinisch probleem 'bovenbuikklachten': amylase, bilirubine, LDH en alkalische fosfatase.

Beschouwing

Wij onderzochten met een RCT in een grote huisartsenpopulatie in vijf regio's in Nederland het DTO met als doel het diagnostisch aanvraagdgedrag van huisartsen te verbeteren. De vermindering in aantallen testaanvragen kan opgevat worden als kwaliteitsverbetering omdat die daling overeenkomt met aanbevelingen uit de NHG-Standaarden. Het verschil in effect tussen de twee armen, de A- en B-klinische problemen, is intrigerend. Beide armen werden zo samengesteld dat er zo weinig mogelijk interferentie tussen de problemen van de verschillende armen kon optreden. Het feit dat in arm II geen significante daling optrad, heeft mogelijk te maken met het kwantitatief belangrijkste probleem in arm II: vage klachten/moeheid. Tijdens de desbetreffende DTO-bijeenkomst werden de richtlijnen uit de NHG-Standaard Bloedonderzoek behandeld. Er bestaan voldoende aanwijzingen dat huisartsen deze standaard moeilijk implementeerbaar vinden.^{21,22} Datzelfde geldt bijvoorbeeld ook voor het onderwerp degeneratieve gewrichtsafwijkingen. De visie uit de standaarden hierover dat röntgenfoto's in het algemeen weinig zinvol zijn voor het beleid is aan patiënten vaak moeilijk uit te leggen en sluit ook niet aan op het diagnostisch beleid in de tweede lijn.

ten opzichte van de totale variantie tussen clusters. Als het aanvraagdgedrag binnen een HAGRO net zoveel verschilt als het aanvraagdgedrag tussen HAGRO's, dan zal de ICC nul zijn. Bij de gegevens uit de nulmeting bleek de ICC voor de tests van blok A en blok B respectievelijk 0,10 en 0,12. Deze ICC betekende dat huisartsen wat hun aanvraagdgedrag betreft binnen hun HAGRO iets meer op elkaar leken dan alle huisartsen onderling. Niettemin had deze clustering dermate weinig invloed op de effecten met de multilevelanalyses dat covariantieanalyses op individueel huisartsenniveau toch mogelijk bleken.

Resultaten

Zesentwintig HAGRO's met 174 huisartsen bleken na een eerste verzoek van de medisch coördinator bereid deel te nemen. Na randomisatie bestonden beide armen uit 13 HAGRO's (figuur). In tabel 3 staan de kenmerken van de huisartsenpopulatie. De gemiddelde groeps grootte en de ervaring met intercollegiale toetsing verschilden niet tussen beide armen. Verder was er bij de nulmeting geen verschil in het gemiddeld aantal tests per huisarts.

Tabel 3 Individuele kenmerken van de onderzoekspopulatie

Individuele kenmerken	Arm I	(SD)	Arm II	(SD)
Leeftijd	46,2	(6,6)	45,8	(5,4)
Aantal vrouwelijke huisartsen	14		15	
Praktijk grootte	2587	(641)	2637	(519)
Percentage patiënten boven de 65 jaar	15	(6,8)	13	(7,1)
Parttime factor	0,91	(0,15)	0,91	(0,16)
Aantal huisartsen werkzaam als solist	43		48	
Aantal geautomatiseerde praktijken	66		61	

Tabel 4 Effecten van het DTO op de aantallen tests (SD) per huisarts per halfjaar

	Arm I (DTO op A)					Arm II (controle)							
	Nulmeting		Follow-up	verschil (%)		Nulmeting		Follow-up	verschil (%)	β^*	standaard-fout β	95%-BI	
Totaal aantal tests arm I:	478	(309)	422	(234)	-12	507	(293)	503	(281)	0	- 67	19	-104 - -30
- cardiovasculair	293	(189)	276	(157)	-6	290	(182)	302	(184)	+4	- 35	13	-61 - -10
- bovenbuikklachten	165	(125)	128	(82)	-22	192	(128)	174	(114)	-9	- 28	9	-45 - -10
- onderbuikklachten	20	(20)	18	(19)	-10	25	(25)	27	(29)	+8	- 5	2	-9 - -1
	Arm I (controle)					Arm II (DTO op B)							
Totaal aantal tests arm II:	640	(394)	624	(357)	-3	724	(386)	664	(356)	-8	-28	23	-74 - 14
- COPD/astma	39	(31)	31	(25)	-20	53	(27)	38	(19)	-28	-1	2	-5 - 3
- vage klachten/moeheid	548	(340)	544	(310)	0	599	(340)	568	(321)	-5	-19	21	-61 - 22
- degeneratieve gewrichtsafwijkingen	54	(38)	49	(36)	-9	72	(43)	58	(37)	-19	-3	4	-10 - 4

* β = interventie-effect, de totale verandering tussen nulmeting en follow-up van het gemiddeld aantal tests in de interventiearm minus de totale verandering tussen nulmeting en follow-up van het gemiddeld aantal tests in de controlearm.
De aantallen zijn gecorrigeerd voor de nulmeting en de regio, zodat simpel op- en aftrekken van de cijfers uit de kolommen niet het eindresultaat oplevert.

Tabel 5 Effecten van het DTO op de volgens richtlijnen 'irrationele' tests (SD) per huisarts per halfjaar

Irrationele tests	Arm I (DTO op A)				Arm II (controle)				β^*	standaard- fout β	95%-BI
	Nulmeting	Follow-up	verschil (%)		Nulmeting	Follow-up	verschil (%)				
Irrationele tests uit arm I:	63 (75)	45 (41)			66 (55)	63 (56)			-16	4,8	-27 - -7
ureum	8,7 (19)	7,2 (15)			6,3 (7,2)	6,6 (8,3)			-1	1,3	-4 - -2
SGOT	7,7 (11)	5,5 (7,7)			8,3 (13)	7,5 (14)			-2	1,4	-5 - -1
LDH	13 (27)	8,8 (16)			12 (20)	11 (18)			-3	1,5	-6 - -1
amylase	5,3 (13)	3,6 (6,9)			3,4 (4,9)	4,5 (10)			-2	1,1	-4 - -0,1
alkalische fosfatase	11 (25)	7,0 (11)			9,3 (13)	9,0 (15)			-3	1,5	-6 - -0,3
bilirubine	20 (27)	15 (19)			31 (43)	27 (35)			-6	2,6	-11 - -0,3
Irrationele tests uit arm II:	Arm I (controle)				Arm II (DTO op B)				β^*	standaard- fout β	95%-BI
	Nulmeting	Follow-up	verschil (%)		Nulmeting	Follow-up	verschil (%)				
totaal IgE	134 (81)	126 (74)			163 (89)	138 (74)			-8	5,0	-18 - 2
leukocytenaantal	3,6 (5,3)	2,8 (4,7)			3,0 (5,3)	1,5 (2,7)			-1	0,42	-1 - 1
totaal aantal röntgenfoto's gewrichten	95 (63)	92 (57)			110 (69)	96 (58)			-6	4,0	-4 - 2
	36 (26)	31 (22)			50 (34)	41 (26)			-1	2,7	-4 - 6

β = interventie effect, de totale verandering tussen nulmeting en follow-up van het gemiddeld aantal tests in de interventiearm minus de totale verandering tussen nulmeting en follow-up van het gemiddeld aantal tests in de controlearm.
De aantallen zijn gecorrigeerd voor de nulmeting en de regio, zodat simpel op- en aftrekken van de cijfers uit de kolommen niet het eindresultaat oplevert.

Beperkingen

Het is mogelijk dat alleen gemotiveerde, goed functionerende HAGRO's wilden meedoen aan het onderzoek en het gevonden effect niet generaliseerbaar is. Verder werden alleen kwantitatieve gegevens gebruikt. Het is natuurlijk niet zo dat een daling van het aantal tests synoniem is met kwaliteitsverbetering en 'hoe minder, hoe beter' geldt zeker niet. Een daling van het aantal 'irrationele' aanvragen lijkt echter een aanwijzing voor kwaliteitsverbetering van het aanvraaggedrag. Overigens is er bewijs dat een reductie in aantallen diagnostische tests niet automatisch leidt tot meer verwijzingen of substitutie van zorg.²³ Tot slot is er geen reden te veronderstellen dat de huisartsen uit de twee armen elkaar onderling zouden kunnen beïnvloeden. Huisartsen bespreken hun aanvraaggedrag nauwelijks structureel met collega's en de groepen kwamen uit verschillende regio's.

Het effect van het DTO blijkt niet groter dan de effecten van de bekende Nederlandse onderzoeken (tabel 6). De resultaten zijn natuurlijk niet volledig vergelijkbaar omdat onderzoeksopzetten, analyses en uitkomstmaten niet dezelfde zijn. Dit gegeven en het feit dat niet voor alle klinische problemen significante dalingen optraden, betekent dat het DTO niet de ultieme strategie is om het aanvraaggedrag te verbeteren. Additionele strategieën zoals

het gebruik van beslissingsondersteunende systemen of probleemgeoriënteerde aanvraagformulieren blijven nodig om het aanvraaggedrag op een positieve manier te beïnvloeden.^{10-12,24}

Conclusie

Het DTO blijkt een effectieve manier om het aanvraaggedrag van huisartsen te verbeteren, hoewel langetermijneffecten nog niet duidelijk zijn. Het overleg gaat uit van de dagelijkse praktijk en blijkt ook goed inpasbaar in die praktijk. Huisartsen krijgen in het

Tabel 6 Vergelijking van effecten van Nederlandse onderzoeken

Onderzoek	Jaar	Interventie	Onderzoeksopzet	Effect*	p
Zaat	1989	verandering aanvraagformulier	pre-post	-13	0,00
Van Geldrop	1990	probleemgeoriënteerd aanvraagformulier	pre-post	-6	<0,05
Winkens	1995	persoonlijke feedback	RCT	-20	0,04
Van Wijk	1996	beslissingsondersteunend systeem	RCT	-20	0,003
Verstappen	1998	DTO	RCT	-12	0,001

* Het effect geeft het verschilpercentage in aantallen tests voor en na de interventie.

DTO de mogelijkheid om ieders aanvraaggedrag open te bediscussieren en ervan te leren. Discussies en afspraken worden minder vrijblijvend wanneer zij over transparante gegevens kunnen discussiëren, tot een HAGRO-consensus over de richtlijnen kunnen komen en daarna problemen kunnen bespreken die bij een eventuele implementatie van ander aanvraaggedrag verwacht kunnen worden. Daarom kan de methode ook interessant zijn voor andere gebieden, zoals verwijs- of prescriptiebeleid. Verdere invoering zou dan ook toe te juichen zijn, zoals momenteel gebeurt vanuit diagnostische centra in verschillende regio's in Zuid-Nederland.

Dankwoord

De auteurs danken het College voor Zorgverzekeringen voor de subsidie waarmee dit onderzoek uitgevoerd kon worden.

Literatuur

- 1 Leurquin P, Van Casteren V, De Maeseneer J. Use of blood tests in general practice: a collaborative study in eight European countries. Eurosentinel Study Group. Br J Gen Pract 1995;45:21-5.
- 2 Ayanian JZ, Berwick DM. Do physicians have a bias toward action? A classic study revisited. Medical Decision Making 1991;11:154-8.
- 3 Kristiansen IS, Hjortdahl P. The general practitioner and laboratory utilization: why does it vary? Fam Pract 1992;9:22-7.
- 4 Holtgrave DR, Lawler F, Spann SJ. Physicians' risk attitudes, laboratory usage, and referral decisions: the case of an academic family practice center. Med Decis Making 1991;11:125-30.
- 5 Zaat JO, Van Eijk JT. General practitioners uncertainty, risk preference, and use of laboratory tests. Med Care 1992;30:846-54.
- 6 Wong ET. Improving laboratory testing: can we get physicians to focus on outcome? Clinical Chemistry 1995;41:1241-7.
- 7 Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Medicine. Communicating statistical information. Science 2000;290:2261-2.
- 8 McDonald IG, Daly J, Jelinek VM, Panetta F, Gutman JM. Opening Pandora's box: the unpredictability of reassurance by a normal test result. BMJ 1996;313:329-32.
- 9 Little P, Cantrell T, Roberts L, Chapman J, Langridge J, Pickering R. Why do GPs perform investigations? The medical and social agendas in arranging back X-rays. Fam Pract 1998;15:264-5.
- 10 Zaat JOM, Van Eijk JT, Bonte HA. Mag het ook een testje minder? De invloed van een beperking van het aanvraagformulier voor laboratoriumonderzoek. Huisarts Wet 1991;34:72-7.

- 11 Geldrop W, Lucassen PBLJ, Smithuis LOMJ. Een probleemgeoriënteerd aanvraagformulier voor laboratoriumonderzoek. Effecten op het aanvraaggedrag van huisartsen. Huisarts Wet 1992;35:192-6.
- 12 Van Wijk MA, Van der Lei J, Mosseveld M, Bohnen AM, Van Bommel JH. De invloed van computer ondersteunende besluitvorming op het aanvragen van bloedonderzoek door huisartsen. Huisarts Wet 2001;44:331-6.
- 13 Winkens RA, Pop P, Bugter Maessen AM, Grol RP, Kester AD, Beusmans GH, et al. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. Lancet 1995;345:498-502.
- 14 Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet 1993;342:1317-22.
- 15 Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. Can Med Ass Journal 1995;153:1423-31.
- 16 Solomon DH, Hashimoto H, Daltroy L, Liang MH. Techniques to improve physicians' use of diagnostic tests. A new conceptual framework. JAMA 1998;280:2020-7.
- 17 Grol R. Peer review in primary care. Qual Assur Health Care 1990;2:219-26.
- 18 Mittman B, Tones X, Jacobson P. Implementing clinical practice guidelines: social influence strategies and practitioner behavior change. Qual Rev Bull 1992;18:413-22.
- 19 Dijkmans JP, Bentum V STB, Keyzer J, Harms LM. Feedback aan huisartsengroepen vanuit een diagnostisch centrum. Med Contact 1995;46:1473-5.
- 20 Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow-up measurements. BMJ 2001;323:1123-4.
- 21 Van der Weijden T, Van Bokhoven MA, Dinant GJ, Van Hasselt CM, Grol RP. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. Br J Gen Pract 2002;52:974-80.
- 22 Van der Weijden T, Van Velsen M, Dinant GJ, Van Hasselt CM, Grol R, Van Bokhoven MA, et al. Unexplained complaints in general practice: prevalence, patients' expectations, and professionals' test-ordering behavior. Designing a quality improvement intervention: a systematic approach. Med Decis Making 2003;23:226-31.
- 23 Winkens RA, Grol RP, Beusmans GH, Kester AD, Knottnerus JA, Pop P. Does a reduction in general practitioners' use of diagnostic tests lead to more hospital referrals? Br J Gen Pr 1995;45:289-92.
- 24 Eccles M, Steen N, Grimshaw J, Thomas L, McNamee P, Soutter J, et al. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. Lancet 2001;357:1406-9.

Hoe kijken huisartsen naar rode ogen?

Niet zoals in de NHG-Standaard wordt aanbevolen

Teun Spies, Henk Mokkink, Pieter de Vries Robbé, Richard Grol

Inleiding

Huisartsen zien in een normpraktijk wekelijks 2 tot 3 keer een patiënten met een rood oog (incidentie 40/1000).¹ Slechts een klein deel van de gevallen betreft een ernstige aandoening, zoals keratitis (incidentie 2,3), iridocyclitis (incidentie 0,6) of glaucoom

(incidentie 0,4).^{1,2} Of er sprake is van een ernstige aandoening, beoordelen huisartsen op grond van het klinische beeld. Ze stellen vragen, inspecteren het oog globaal en doen eventueel een aantal meer gerichte oogonderzoeken, zoals visusbepaling, beoordeling van de pupilreacties, van de voorste oogkamer bij zij-