

Kwaliteit van huisartsgeneeskundig handelen moeilijk te meten

Een literatuuroverzicht over 1997-2003

J Braspenning, AM Schiere, M van Roosmalen, H Mookink, RPTM Grol

Inleiding

Hoewel in de samenleving de behoefte aan informatie over de kwaliteit van de patiëntenzorg groeit, is de informatie over de zorg in de Nederlandse huisartsenpraktijk schaars. Nu is het nog maar de vraag of de kwaliteit van zorg te meten is. We hebben weliswaar enkele meetinstrumenten voor de praktijkvoering,¹ de communicatie tussen arts en patiënt²⁻⁴ en het oordeel van de patiënt

Samenvatting

Braspenning J, Schiere AM, Van Roosmalen M, Mookink H, Grol RPTM. Kwaliteit van huisartsgeneeskundig handelen moeilijk te meten. Een literatuuroverzicht over 1997-2003. *Huisarts Wet* 2004;47(4):184-7.

Doel Het evalueren van het feitelijk medisch handelen in de Nederlandse huisartsenpraktijk.

Methode Systematisch literatuuroverzicht over de periode 1997-2003.

Resultaat Eenendertig onderzoeken voldeden aan de inclusiecriteria. De geboden informatie ging over een beperkt deel van het huisartsgeneeskundig handelen; het merendeel chronische aandoeningen (77,8%), waaronder astma en COPD (n=6), diabetes (n=5) en risicomangement van cardiovasculaire aandoeningen (n=5). Niet alleen verschilde de manier van meten tussen de diverse aandoeningen, ook werd het handelen van de huisartsen bij dezelfde aandoening in de verschillende onderzoeken telkens anders gemeten. De overeenstemming tussen het feitelijk medisch handelen en de aanbevelingen lag gemiddeld op 62,4%, maar varieerde sterk (bereik 14,3-90,1%).

Conclusie De beschikbare onderzoeken bieden onvoldoende aanknopingspunten om het feitelijk medisch handelen van huisartsen valide en betrouwbaar te evalueren. Om het feitelijk medisch handelen transparant te maken is een systematische aanpak noodzakelijk waarbij er voor meer aandoeningen indicatoren moeten worden ontwikkeld.

UMC St Radboud, Kwaliteit van Zorg, WOK, Postbus 9101, 6500 HB Nijmegen; dr. J. Braspenning, onderzoekspsycholoog; A.M. Schiere, socioloog; M. van Roosmalen, epidemioloog; dr. H. Mookink, socioloog; prof.dr. R. Grol, hoogleraar Kwaliteit van Zorg.
Correspondentie: J.Braspenning@wok.umcn.nl
Mogelijke belangenverstrengeling: niets aangegeven.

over de zorg,⁵⁻⁸ maar over valide en betrouwbare meetinstrumenten om het feitelijk medisch handelen te meten is nog weinig bekend.

Wij hebben een literatuuroverzicht uitgevoerd naar meetinstrumenten om het medisch-technisch handelen in de huisartsenpraktijk te bepalen en om na te gaan of dit handelen in overeenstemming is met de geformuleerde kwaliteitscriteria, zoals NHG-Standaarden. In dit overzicht beschrijven wij de 'kwaliteit' van het *feitelijk* medisch handelen op basis van een literatuuroverzicht over een tijdsvak van bijna zeven jaar: 1997-2003. Wij onderzochten over welke aandoeningen en verrichtingen informatie bekend is over het feitelijk handelen en of dat in overeenstemming is met achterliggende aanbevelingen. Daarbij gingen wij na of een verschil in handelen gerelateerd was aan het gehanteerde kwaliteitscriterium (NHG-Standaard of andere aanbeveling); de aandoening (chronisch of niet-chronisch) en de verrichting, waarbij wij een onderscheid maakten naar preventie, diagnostiek, educatie, prescriptie, verwijzing of follow-up.

Methode

Zoekstrategie en selectie van onderzoeken

Aan de hand van zoektermen over de huisartsgeneeskunde en kwaliteit van zorg hebben wij in 2 elektronische databases (Medline en Picarta) gezocht naar Engels- en Nederlandstalige artikelen (gerefereerde tijdschriften) gepubliceerd in de periode 1997-2003 (tot november 2003). In *tabel 1* staan de gebruikte MESH-termen. Dit leverde 2193 treffers op, waarvan 121 betrekking hadden op Nederlandse huisartsen. Wij hebben de artikelen beoordeeld op de volgende inclusiecriteria:

- ▶ feitelijk handelen gemeten (geen vignetonderzoeken of opinies);

Tabel 1 Gebruikte zoektermen

MESH-termen

Engelstalig: family practice/general practice AND quality of care, audit, clinical competence, feedback, guideline adherence/compliance, healthcare peer review, healthcare quality indicators/assurance, total quality management, quality control of continuous quality improvement.

Nederlandstalig: huisarts(praktijk) EN visitatie, kwaliteit van zorg, deskundigheid(sbevordering), standaarden/richtlijnen, kwaliteitsbevordering, -bewaking, -management, -verbetering, -indicatoren, borging, systeem, gezondheidszorgonderzoek, implementatie, (klinische) (deskundigheids) educatie of interventie.

- ▶ kwaliteitscriterium aanwezig (NHG-Standaard of mening van experts);
- ▶ methodologische kwaliteit van het onderzoek (minimaal 5 praktijken, normale gang van zaken, gepresenteerde resultaten zijn herkenbaar af te leiden uit de analyses).

Beoordelingsinstrument

De artikelen zijn beoordeeld op vooraf ontworpen scoringsformulieren waarop naast de inclusiecriteria ruimte was voor een inhoudelijke beschrijving aan de hand van de volgende punten: onderzoeksonderwerp (aandoening en verrichting), onderzoekspopulatie (aantal huisartsen en patiënten) en de resultaten, dat wil zeggen in welke mate het feitelijk medisch handelen in overeenstemming was met het kwaliteitscriterium. Het handelen uitgedrukt in maat en getal wordt een indicator genoemd. Analoog aan de indeling van Donabedian hielden wij een onderscheid aan tussen structuur-, proces- of uitkomstindicatoren.⁹ Voor de diabeteszorg bijvoorbeeld, is het houden van een diabetespreekuur een structuurindicator, het wel of niet uitvoeren van voetonderzoek bij de jaarlijkse controle is een procesindicator en de HbA1c-waarde een uitkomstindicator (een *proxy* voor het aantal complicaties gedurende een bepaalde tijdsperiode). Als er sprake was van een interventieonderzoek, dan zijn alleen de gegevens van de voormeting gebruikt. De scoringsformulieren zijn door twee personen (JB, AMS of MH) afzonderlijk ingevuld. Eventuele meningsverschillen werden door discussie opgelost en indien nodig werd een derde persoon geconsulteerd.

Analyse

De kwaliteit van het medisch handelen beoordeelden we door het in de artikelen beschreven en gemeten handelen af te zetten tegen het gehanteerde kwaliteitscriterium en uit te drukken in een percentage. Als in een artikel meer dan één meting is beschreven dan hebben wij de percentages opgeteld en gemiddeld. Het gebruikte overall percentage is steeds zodanig beschreven dat de hoogte van het percentage de mate van overeenstemming weerspiegelt tussen handelen en aanbeveling. Er is een onderscheid gemaakt naar structuur-, proces- en uitkomstindicatoren. Als in één onderzoek meerdere typen metingen (structuur, proces of uitkomst) zijn verricht, dan is/zijn voor het gemiddelde alleen de procesindicator(en) meegenomen met het oog op de mogelijke afhankelijkheid tussen de verschillende typen metingen.

Met een variantieanalyse toetsten we of de volgende determinanten van invloed zijn op de mate van handelen volgens de aanbeveling: het type kwaliteitscriterium, de soort aandoening en het type verrichting. Het aantal deelnemende huisartsen(praktijken) en het aantal patiënten dan wel beslissingen zijn als covariaat in het model meegenomen. Er is eenzijdig getoetst, waarbij we een significantieniveau van $p < 0,05$ hanteerden.

Resultaten

Inclusiecriteria

Van de 121 beoordeelde artikelen voldeden 31 aan de inclusiecri-

Wat is bekend?

- ▶ Onze kennis over de kwaliteit van de geleverde zorg in de huisartsenpraktijk is beperkt.

Wat is nieuw?

- ▶ In de afgelopen 7 jaar verschenen 31 Nederlandse onderzoeken waarin het feitelijk handelen is getoetst aan richtlijnen. De onderzoeken gaan vooral over een beperkt aantal chronische aandoeningen.
- ▶ Betrouwbare en valide meetinstrumenten om het huisartsge-neeskundig handelen transparant te maken zijn schaars.^a

teria; de lijst van geselecteerde artikelen staat op de website: <http://nhg.artsennet.nl>. De grootste uitval ontstond omdat er geen feitelijk handelen werd beschreven ($n=77$). Nog eens 5 artikelen vielen af omdat dezelfde gegevens in meerdere artikelen werden gepresenteerd (éénmaal opgenomen). Vier artikelen voldeden niet aan onze methodologische voorwaarden en in 4 artikelen werd geen kwaliteitscriterium genoemd.

Onderwerpen

De 31 onderzoeken gaan over een beperkt aantal aandoeningen: 6 over astma en COPD bij volwassenen, 5 over diabetes, 5 over risicomangement bij hart- en vaatziekten; één over cholesterol en één over hypertensie (tabel 2, website: <http://nhg.artsennet.nl>). Twee onderzoeken beschrijven cervixscreening en één influenza-vaccinatie. De andere onderwerpen betreffen dementie, darmklachten, HIV-infectie, (dreigende) miskraam, verloskunde, otitis media acuta, en het achterhalen van de reden van het consult. Daarnaast worden soms meerdere onderwerpen in één artikel beschreven. In één artikel worden 10 onderwerpen behandeld, waaronder otitis media acuta, diabetes en hypertensie; één artikel gaat over diverse testaanvragen bij 16 aandoeningen en één artikel gaat in op 251 voorgeschreven middelen (medicatie).

Bij 16 van de 31 onderzoeken ging het om één of meerdere NHG-Standaarden. In 21 onderzoeken (77,8%) was er sprake van een chronische aandoening. De onderzochte verrichtingen liepen sterk uiteen, maar zowel preventie, diagnostiek, educatie, prescriptie, verwijzing als follow-up kwamen voor; preventie inclusief de diabeteszorg en de cardiovasculaire aandoeningen (13 onderzoeken) en prescriptie (11 onderzoeken) nog het meest. Per type verrichting werd soms één handeling gemeten, maar vaak ging het om meerdere (het maximum lag op 44).

Overeenstemming tussen handelen en aanbevelingen

De gedane metingen betroffen doorgaans procesindicatoren (22 onderzoeken), in vier onderzoeken was sprake van alleen structuurindicatoren en in één onderzoek ging het alleen om uitkomstindicatoren. Procesindicatoren zijn in één onderzoek gecombineerd met structuurindicatoren, in twee onderzoeken met uitkomstindicatoren en in één onderzoek met beide. De relatie tussen het type indicator en de kwaliteit van het handelen kon niet getoetst worden omdat het aantal onderzoeken met struc-

Tabel 2 en de reviewlijst staan op de website:
<http://nhg.artsennet.nl/henw>

tuur- en uitkomstindicatoren te gering was: respectievelijk 6 en 4. De overeenstemming tussen handelen en aanbeveling varieerde voor de structuurindicatoren van 13,7% tot 57,4%, voor de procesindicatoren van 14,3% tot 90,1% en voor de uitkomstindicatoren van 38,9% tot 70,9%.

Op itemniveau varieerde het handelen nog sterker met een minimum van 10,5% (registratie preventieve activiteiten en teamoverleg) en een maximum van 96,5% (geen testaanvragen bij darmklachten). Het gemiddelde percentage voor de procesindicatoren was 62,4% (SD 21,1).

Determinanten van compliance aan aanbevelingen

Voor het determinantenonderzoek hebben we ons beperkt tot de procesindicatoren en eerst gekeken naar de invloed van het kwaliteitscriterium (28 onderzoeken). Huisartsen volgden in deze onderzoeken de NHG-Standaarden niet beter dan de adviezen van experts: 58,5% van de aanbevelingen uit de NHG-Standaarden werd gevolgd (tabel 3) en 67,6% van de aanbevelingen van de experts (meestal onderdeel van het onderzoeksteam).

Vervolgens hebben we in één model het soort aandoening en het type verrichting getoetst. Voor de aandoeningen hebben we op basis van de beschikbare onderzoeken gekozen voor een onderscheid in chronische en niet-chronische aandoeningen en in wel of geen preventietaken, waarbij de procesindicatoren in de diabetesonderzoeken zijn geduid als preventietaken (26 onderzoeken, zie tabel 3). Voor chronische aandoeningen is de overeenstemming tussen handelen en aanbeveling iets lager dan voor niet-chronische aandoeningen (58,7% versus 70,3%), maar dit verschil was niet significant. We vonden geen significant verschil tussen aanbevelingen voor preventieactiviteiten (64,8%) en de andere verrichtingen (59,3%). De aanbevelingen worden minder goed

gevolgd als het om voorschrijven van medicatie gaat (53,3%) dan wanneer het andere verrichtingen betreft (67,1%), maar ook dit verschil voldoet niet aan het gestelde significantieniveau.

Beschouwing

Onze resultaten laten vooral zien dat er nog weinig harde gegevens zijn over de overeenstemming tussen het feitelijk handelen en allerlei aanbevelingen voor het medisch handelen door de huisarts. In totaal hebben we voor een periode van bijna 7 jaar (1997-2003) 31 onderzoeken gevonden. Het ging daarbij vaak om astma en COPD, diabetes en risicomangement van hart- en vaatziekten, maar het onderzochte handelen voor deze aandoeningen liep sterk uiteen. Ook voor de andere aandoeningen zien we een breed scala aan onderzochte verrichtingen. Als we naar de mate van overeenstemming tussen handelen en de verschillende aanbevelingen kijken, dan is het meest opvallende resultaat de enorme variatie. De variatie is eigenlijk nog groter dan beschreven, omdat wij ervoor gekozen hebben de verrichtingen per onderzoek te sommeren en te middelen.

Wij hebben geprobeerd om de kwaliteit van het feitelijk medisch handelen vast te stellen. Overal was het percentage handelen door huisartsen volgens de richtlijnen voor de procesindicatoren 62,4%. In tabel 2 wordt echter bij een drietal onderzoeken aangegeven dat de onderzochte praktijken al gedurende langere tijd in het meten van medisch handelen werden geschoold. Het genoemde percentage is dus eerder een overschatting dan een onderschatting. De determinantenanalyse leverde geen significante resultaten op. De aantallen onderzoeken zijn echter ook erg klein. Hoe gevarieerder de onderzoeken zijn, des te meer hebben we er eigenlijk nodig om betrouwbare uitspraken te kunnen doen. Het lijkt erop dat het voorschrijven minder conform de richtlijnen wordt uitgevoerd dan andere verrichtingen, met name bij chronische aandoeningen. NHG-Standaarden worden blijkbaar niet beter gevolgd dan andere richtlijnen, maar het gegeven dat de aanbevelingen van experts vaak afkomstig waren van het onderzoeksteam kan hierin een factor van betekenis zijn geweest. Een deel van de variatie zou nog te maken kunnen hebben met het registeren in het medische dossier. De meeste onderzoeken (n=23) hebben voor hun gegevens gebruikgemaakt van het (elektronisch) medisch dossier en in de andere acht onderzoeken is gebruikgemaakt van registratielijsten. Zoals bekend zijn niet alle gegevens terug te vinden in een medisch dossier.

Onze conclusie komt overeen met die van vergelijkbare overzichtsartikelen uit de Verenigde Staten over de eerste- en tweedelijnsgezondheidszorg en uit Groot-Brittannië over de huisartsenpraktijk.^{10,11} De onderwerpen hebben ook daar voor het merendeel betrekking op chronische aandoeningen. De geconstateerde variatie in handelen leidde daar tot een kwaliteitssysteem met feedback over de geleverde zorg en discussies over aanpassingen van de onderzoeks-, klinische en de politieke agenda. Voordat er strategieën ter verbetering ontwikkeld kunnen worden, moet volgens ons eerst een meer basale vraag beantwoord

Tabel 3 De invloed van het gehanteerde kwaliteitscriterium, soort van aandoening en type van activiteit (preventie, prescriptie) op het gemiddelde handelen volgens de aanbevelingen

| | gemiddeld percentage | spreiding (SD) | p | aantal onderzoeken |
|---|----------------------|----------------|-------|--------------------|
| <i>Kwaliteitscriterium NHG-Standaard:</i> | | | | |
| - nee | 67,6 | 15,9 | 0,135 | 12 |
| - ja | 58,5 | 24,1 | | |
| <i>Chronische aandoening:</i> | | | | |
| - nee | 70,3 | 27,8 | 0,431 | 5 |
| - ja | 58,7 | 20,8 | | |
| <i>Preventie-activiteiten:</i> | | | | |
| - nee | 59,3 | 20,1 | 0,258 | 15 |
| - ja | 64,8 | 26,4 | | |
| <i>Prescriptie-indicatoren:</i> | | | | |
| - ja | 67,1 | 22,3 | 0,058 | 14 |
| - nee | 53,3 | 20,6 | | |

Abstract

Braspenning J, Schiere AM, Van Roosmalen M, Mokkink H, Grol RPTM. Quality of care in general practice is difficult to measure. A review 1997-2003. *Huisarts Wet* 2004;47(4):184-7.

Background Little is known about the quality of clinical care provided in Dutch general practices. The aim of the study was to assess actual clinical performance in Dutch general practices.

Method A systematic review of published studies in a seven-year period 1997-2003.

Results Thirty-one studies could be included. The majority of the studies were on chronic diseases (77.8%); asthma and COPD (n= 6), diabetes (n=5), cardiovascular risk management (n=5), and one study each on hypercholesterolemia, hypertension, dementia, bowel complaints, and HIV-infection. Furthermore, two studies were found on cervical cancer screening, and one on influenza vaccination; the others were on, acute otitis media, (threatened) miscarriage, child birth and getting to know the reason for consultation. One study was on 10 diseases, among them acute otitis media, diabetes, and hypertension. One study was on test ordering for 16 conditions and one on 251 different prescriptions. The agreement between performance and recommendation varied largely; the overall percentage was 62.4%.

Discussion The study showed that since the development of the national guidelines in the early nineties, the amount of studies assessing the clinical performance in general practice is small and that the published studies are on a limited set of diseases in which the variance in performance is enormous between the different diseases, but also for the same disease. This brings us to the question of the value of current performance measurement, also because the studies don't describe a systematic methodology for constructing the indicators and their validity and reliability. Transparency on clinical performance in general practice requires a systematic approach to fulfil the demand on a valid and reliable set of indicators and to create a more comprehensive picture of the performance in general practice.

worden, namelijk in welke mate daadwerkelijk op valide wijze de kwaliteit van zorg is gemeten.

Maar zelden troffen wij een systematische afweging van geschikte indicatoren aan. Over de validiteit en betrouwbaarheid van de gebruikte indicatoren werd in geen enkel onderzoek gerapporteerd. Zijn de indicatoren in de onderzoeken systematisch afgewogen? Hoe is het mogelijk dat zelfs indicatoren over dezelfde aandoeningen zo verschillend zijn? Hoe komt het dat bij de ene aandoening zoveel meer volgens een aanbeveling wordt gehandeld dan bij een andere aandoening? Veel vragen blijven nog onbeantwoord. We blijken eigenlijk niet goed te weten hoe we de kwaliteit van het huisartsgeneeskundig handelen moeten meten en welke instrumenten daarbij valide en betrouwbaar zijn.

Voor een kwaliteitssysteem in de Nederlandse huisartsenpraktijk hebben we valide en betrouwbare kwaliteitsindicatoren nodig. Deze indicatoren zouden een breed palet van zorg (aandoeningen en verrichtingen) moeten beslaan. In het kader van de tweede Nationale Studie zijn binnen het Landelijk Informatie Netwerk Huisartsenzorg (LINH) gegevens verzameld voor 139 kwaliteitsin-

dicatoren die op systematische wijze zijn geconstrueerd en zijn gebaseerd op 70 NHG-Standaarden.¹²⁻¹⁴ Naast indicatoren op het medisch-technische gebied zou ook aandacht moeten zijn voor andere aspecten van kwaliteit van zorg, zoals de praktijkorganisatie, de arts-patiëntcommunicatie en het oordeel van de patiënt over alle aspecten. Veel meetinstrumenten zijn al ontwikkeld,¹⁻⁸ maar de meeste zijn nogal arbeidsintensief als het gaat om het verkrijgen van gegevens. Transparantie van kwaliteit en doelmatigheid van zorg in de huisartsenpraktijk vereisen ook een realistische methode van gegevensverzameling. Alleen dan kan de zorg getoetst worden tegen duidelijke kwaliteitscriteria, tussen praktijken en in een tijdsbestek

Dankbetuiging

Met dank aan Mirjam Harmsen die met daadkracht heeft bijgedragen aan de aanvulling op het literatuuroverzicht.

Literatuur

- 1 Van den Hombergh P, Grol R, Van den Hoogen H, Van den Bosch WJ. Practice visits as a tool in quality improvement: acceptance and feasibility. *Qual Health Care* 1999;8:167-71.
- 2 Bensing J. Doctor-patiënt communication and the quality of care. *Soc Sci Med* 1991;32:1301-10.
- 3 Ram P, Grol R, Rethans JJ, Schouten B, Van der Vleuten C, Kester A. Assessment of general practitioners by video observation of communicative and medical performance in daily practice: issues of validity, reliability and feasibility. *Med Educ* 1999;33:447-54.
- 4 Van den Brink-Muinen A, Verhaak PF, Bensing JM, Bahrs O, Deveugele M, Gask L, et al. Communication in general practice: differences between European countries. *Fam Pract* 2003;20:478-85.
- 5 Wensing M, Van der Vleuten C, Grol R, Felling A. The reliability of patients' judgements of care in general practice: how many questions and patients are needed? *Qual Health Care* 1997;6:80-5.
- 6 Jung HP, Van Horne F, Wensing M, Hearnshaw H, Grol R. Which aspects of general practitioners' behaviour determine patients' evaluations of care? *Soc Sci Med* 1998;47:1077-87.
- 7 Grol R, Wensing M, Mainz J, Ferreira P, Hearnshaw H, Hjortdahl P, et al. Patients' priorities with respect to general practice care: an international comparison. European Task Force on Patient Evaluations of General Practice (EUROPEP). *Fam Pract* 1999;16:4-11.
- 8 Sixma HJ, Van Campen C, Kerssens JJ, Peters L. Quality of care from the perspective of elderly people: the QUOTE-elderly instrument. *Age Ageing* 2000;29:173-8.
- 9 Donabedian A. Explorations in quality assessment and monitoring (vol 1): the definition of quality and approaches to its assessment. Ann Arbor, Mi: Health Administration Press, 1980.
- 10 Schuster M, McGlynn E, Brook R. How good is the quality of health care in the United States? *Milbank Q* 1998;76:517-63.
- 11 Seddon ME, Marshall MN, Campbell SM, Roland MO. Systematic review of studies of quality of clinical care in general practice in the UK, Australia and New Zealand. *Qual Health Care* 2001;10:152-8.
- 12 www.linh.nl
- 13 Schellevis FG, Westert GP, De Bakker DH, Groenewegen PP, Van der Zee J, Bensing JM. De tweede Nationale Studie naar ziekten en verrichtingen in de huisartsenpraktijk: aanleiding en methoden. *Huisarts Wet* 2003;46:7-12.
- 14 Campbell SM, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002;11:358-64.