



KWALITEIT IN DE HUISARTSGENEESKUNDE

Hoe wordt kwaliteit van huisartsgeneeskundige zorg meetbaar gemaakt?
Een literatuuronderzoek

6

J.H. Rengelink-van der Lee

november 1988

Nederlands instituut voor onderzoek van de eerstelijnsgezondheidszorg (NIVEL) - Postbus 1568 - 3500 BN Utrecht - telefoon: 030-319946

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Rengelink-van der Lee, J.H.

Kwaliteit in de huisartsgeneeskunde : hoe wordt kwaliteit van huis-
artsgeneeskundige zorg meetbaar gemaakt; een literatuuronderzoek /
J.H. Rengelink-van der Lee. -

Utrecht : Nederlands Instituut voor Onderzoek van de Eerstelijnsge-
zondheidszorg (NIVEL)

Met lit. opg.

ISBN 90-6905-082-X

SISO 604.2 UDC 614.2

Trefw.: huisartsgeneeskunde.

Tekstverwerking: Simone Kerkhove

Vormgeving : Mieke Cornelius

VOORWOORD

Deze literatuurstudie maakt deel uit van een project, waarin het meten van de kwaliteit van de eerstelijnszorg onderwerp van studie is. In het kader van dit project verschijnen drie deelrapporten over respectievelijk de huisartsenzorg, de verloskundige zorg en de wijkverpleging.

Parallel daaraan worden in samenwerking met de bibliotheek van het NIVEL bibliografieën samengesteld waarin korte beschrijvingen van de opgenomen publikaties staan. Deze bibliografieën zijn meer omvattend, omdat ze ook beschouwende artikelen over 'quality assessment' bevatten.

Alhoewel is geprobeerd een zo volledig mogelijk overzicht te geven van onderzoeken waarbij kwaliteit gemeten wordt, zijn wij ons ervan bewust dat de literatuur op dit terrein nog omvangrijker is, dan in dit rapport tot uiting komt. Wij zijn zeer erkentelijk voor de tijd en moeite die R. Grol, J.J.E. Rethans, T. van de Rijdt en Chr. van Weel hebben willen steken in het screenen van de literatuurlijst op belangrijke omissies.

D. Wijkkel, projectleider

INHOUD

	pag.
1. INLEIDING	1
1.1. Definiëring van kwaliteit	2
1.2. Begrenzing van het literatuuronderzoek	3
2. GEVOLGDE WERKWIJZE	7
3. DE DIMENSIES VAN MEETMETHODEN VAN KWALITEIT	9
3.1. Kwaliteitsmeting op processniveau	9
3.2. Kwaliteitsmeting op outcomeniveau	18
3.3. Meting van kwaliteit op zowel process- als outcome-niveau	21
4. ASPECTEN VAN MEETMETHODEN	25
4.1. Bronnen van informatie	25
4.2. Criteria	31
5. BESCHOUWING EN AANBEVELINGEN	34
LITERATUUR	37
BIJLAGEN	51

1. INLEIDING

Het begrip 'kwaliteit van medische zorg' is de laatste twintig jaar steeds meer in de belangstelling gekomen. Hoewel er ook al eerder onderzoeken en verhandelingen werden gepubliceerd over dit onderwerp, bijvoorbeeld Codman, 1914 (aangehaald in Kekki, 1983) en Lee en Jones, 1933 (aangehaald in Donabedian, 1966), is met name sinds de jaren zestig een steeds groeiende stroom van publikaties op gang gekomen.

Voor een belangrijk deel heeft dit te maken met de veranderende maatschappelijke verhoudingen. De patiënt neemt niet meer blindelings aan, dat wat de dokter doet goed is, maar vraagt om informatie en inspraak. Als gevolg van deze ontwikkeling leeft er bij steeds meer artsen de behoefte hun handelen te expliciteren en verantwoordelijkheid met de patiënt te delen. In deze zelfde periode heeft zich in de huisartsgeneeskunde een ontwikkeling voorgedaan in de richting van toenemende professionalisering; de huisartsgeneeskunde is een zelfstandige discipline geworden en wil dit graag naar buiten toe tonen. Eén van de uitingen hiervan is de opkomst van de intercollegiale toetsing in de huisartsgeneeskunde; overigens in navolging van wat eerder in ziekenhuizen door specialisten werd ontwikkeld.

Een later opgetreden maatschappelijk verschijnsel dat geleid heeft tot vergroting van de belangstelling voor het begrip kwaliteit, is dat van de kostenbeheersing. Zowel bij artsen als bij patiënten leeft ongerustheid dat kostenbesparingen zullen leiden tot kwaliteitsvermindering. Besparingen worden alleen dan acceptabel geacht, als ze niet automatisch vermindering van kwaliteit tot gevolg hebben.

Daarnaast is zowel bij artsen als bij patiënten het besef groeiende, dat de toename van medische voorzieningen in de laatste decennia niet alleen positieve effecten teweeg heeft gebracht, maar ook negatieve, zoals het probleem van medicalisering. Doordat gebleken is, dat niet alleen de vraag van belang is of er wel voldoende gedaan wordt, maar ook de vraag of er wellicht niet teveel gedaan wordt, verschuift de belangstelling van kwantiteit naar kwaliteit.

Om op betrouwbare wijze na te gaan of activiteiten met als doel kwaliteitsverbetering, zoals intercollegiale toetsing, inderdaad leiden tot verhoging van de kwaliteit en of bezuinigingen mogelijk zijn zonder aantasting van de kwaliteit, zijn gevalideerde methoden nodig

om het niveau van de kwaliteit van zorg vast te kunnen stellen. In de afgelopen jaren is veel gedaan om dit soort methoden te ontwikkelen. Met name in de literatuur uit de Verenigde Staten, maar ook in de Britse en Nederlandse literatuur zijn ontelbare publikaties over dit onderwerp verschenen. Het doel van dit literatuuronderzoek is een overzicht te geven van de belangrijkste meetmethoden om kwaliteit te meten die toepasbaar zijn in de Nederlandse huisartsgeneeskunde; de gevonden meetmethoden zullen bovendien worden beoordeeld op betrouwbaarheid en validiteit. Met 'betrouwbaarheid' bedoelen wij de 'reproduceerbaarheid' = mate van overeenkomst tussen herhaalde metingen of duplo's en met 'validiteit' de mate waarin een methode werkelijk meet wat ze bedoelt te meten (Sturmans, 1982).

Wij richten ons hierbij vooral op het beschouwen van de validiteit van de **meetmethode**. Op de validiteit van het **begrip** 'kwaliteit', dat auteurs hanteren (is wat men wil meten werkelijk 'kwaliteit?'), zal slechts zijdelings worden ingegaan.

1.1. Definiëring van kwaliteit

Over de **definiëring** van het begrip kwaliteit is veel geschreven. Vaak blijft de definitie van kwaliteit vrij vaag, bijvoorbeeld in beleidsstukken. Onder 'kwaliteit' verstaat men dan een soort algemeen begrip van 'wat goed of wenselijk is'. Toch zijn er vele serieuze pogingen ondernomen kwaliteit te definiëren. Eén van de eerste hiervan is uit 1933, afkomstig van Lee en Jones (aangehaald in Donabedian, 1966). In de vorm van acht 'articles of faith' wordt een waarde-oordeel gegeven ten aanzien van een aantal dimensies van medische zorg, zoals dat in die tijd (en grotendeels nog steeds) de normen en waarden van de westerse maatschappij reflecteerde. Door de Nationale Raad voor de Volksgezondheid is in 1986 de 'Discussienota begrippenkader kwaliteit beroepsuitoefening' uitgebracht, waarin een poging wordt gedaan tot definiëring van kwaliteit volgens de huidige Nederlandse maatstaven. Kwaliteit van de beroepsuitoefening wordt hierin onderverdeeld in drie categorieën: kwaliteit van het methodisch-technisch handelen, kwaliteit van de attitude van de beroepsbeoefenaar en kwaliteit van de organisatie van de beroepsuitoefening, waarbinnen min of meer concrete kwaliteitsaspecten onderscheiden kunnen worden.

Het is duidelijk dat 'kwaliteit' een subjectief begrip is, waarvan de invulling afhankelijk is van de tijd waarin, de plaats waar en de persoon door wie het gedefinieerd wordt. Door Donabedian wordt hier-

op uitgebreid ingegaan in zijn boek 'The definition of quality and approaches to its assessment' (Donabedian, 1980). Om een groot aantal verschillende onderzoeken met verschillende doeleinden en vaak matig geëxpliciteerde uitgangspunten in één overzicht samen te brengen, maken wij hierbij gebruik van de zeer veelomvattende definitie die Donabedian (onder andere) geformuleerd heeft:

'kwaliteit is de mate van overeenstemming tussen de werkelijkheid en tevoren vastgestelde criteria', oftewel:

'kwaliteit is de verhouding van werkelijkheid ten opzichte van wenselijkheid'.

Wat men kwaliteit noemt, hangt dus af van wat men als wenselijkheid formuleert.

Om het veelomvattende begrip 'kwaliteit' beter hanteerbaar te maken, kan men een onderverdeling maken naar niveau's of dimensies waarop men kwaliteit kan meten.

De klassieke indeling van het begrip kwaliteit is die van Donabedian in structure, process en outcome (Donabedian, 1966). Wij zullen deze indeling hanteren, omdat dit de meest algemeen geaccepteerde is. Auteurs die melding maken van een theoretische achtergrond bij hun meetmethoden, gaan vrijwel altijd uit van de indeling structure, process, outcome. 'Structure' staat voor de 'voorwaarden' van goede zorg, zoals opleiding van de hulpverleners, aanwezigheid van apparatuur, bereikbaarheid van een voorziening; met 'process' wordt bedoeld datgene wat plaatsvindt in het contact tussen hulpverlener en patiënt. Hieronder vallen bijvoorbeeld gespreksvoering, lichamelijk onderzoek, verwijzen; en met 'outcome' wordt bedoeld het resultaat van de behandeling in termen van gezondheidstoestand, complicaties of satisfactie.

1.2. Begrenzing van het literatuuronderzoek

De hoeveelheid literatuur over het onderwerp 'kwaliteit van geneeskundige zorg' is, zoals gezegd, zeer omvangrijk. In het kader van dit project zou het ondoenlijk zijn volledigheid na te streven. Daarom is bij de opzet van dit literatuuronderzoek een aantal begrenzingen aangebracht. De eerste begrenzing van het literatuuronderzoek wordt al aangegeven door de hierboven geformuleerde doelstelling: '... die toepasbaar zijn in de Nederlandse huisartsgeneeskunde'. Veel van de literatuur over kwaliteitsmeting is afkomstig uit de Verenigde Staten, waar de huisarts, zoals we die in Nederland kennen, niet voorkomt. Wat deze literatuur betreft, is er daarom

vooral gekeken naar 'ambulatory care' en 'family practice', hoewel dit niet altijd vergelijkbaar is met de Nederlandse huisartsgeneeskunde. De toepasbaarheid in de ons bekende situatie is echter een belangrijk aandachtspunt bij de beschouwing van de literatuur. Als tweede begrenzing willen wij stellen, dat dit literatuuronderzoek zich voornamelijk concentreert op beschrijvingen van onderzoek waarbij het **meten** van kwaliteit centraal staat (Eng.: quality assessment)*. Hiermee sluiten wij een grote hoeveelheid literatuur uit, die zich bezighoudt met begrippen als kwaliteitsbevordering of kwaliteitsbewaking (Eng.: quality assurance).

De algemene gang van zaken bij activiteiten van quality assurance is in grote lijnen als volgt: vaststellen op welke onderwerpen men zich richt, vaststellen wat wenselijk is, meten wat er in werkelijkheid gebeurt, de werkelijkheid toetsen aan de wenselijkheid, feedback van dit resultaat aan de betrokkenen, eventueel bijstellen van de vaststelling van de wenselijkheid en weer opnieuw meten, etcetera.

Het meten van kwaliteit is niet de belangrijkste activiteit van dergelijke programma's; het ontwikkelingskarakter, het bevorderen van de kwaliteit van het handelen van individuele hulpverleners staat voorop. De definiëring van het begrip kwaliteit draagt in veel quality assurance-projecten een ad hoc karakter. Kwaliteit wordt gedefinieerd in termen van hoe hulpverleners zouden willen werken, of de definitie komt alleen tot uiting in het vaststellen van de wenselijkheid voor het beperkte onderwerp waarop het project zich op dat moment toevallig richt. Om deze verschillende redenen is bij dit literatuuronderzoek de meeste literatuur over quality assurance buiten beschouwing gelaten. Wel is die literatuur opgenomen waarin een duidelijke beschrijving werd gegeven van een systematische kwaliteitsmeting.

Binnen het ruime kwaliteitsbegrip volgens de definitie kan men een aantal dimensies onderscheiden, die allemaal een onderdeel vormen van het begrip kwaliteit en die in de literatuur vaak afzonderlijk gepresenteerd worden als equivalent van 'kwaliteit'. Door Williamson wordt een drietal deelaspecten van kwaliteit gedefinieerd, te weten efficacy, effectiveness en efficiency, respectievelijk werkzaamheid, doeltreffendheid en doelmatigheid (Williamson, 1978). Wanneer in het

* Dit wil echter niet zeggen dat het meten van kwaliteit het uiteindelijke doel is. Kwaliteit wordt gemeten om een bepaalde hypothese te toetsen, en niet als een op zichzelf staande activiteit. In dit literatuuronderzoek staat het meten van kwaliteit op zich echter in het middelpunt.

huidige literatuuronderzoek gesproken wordt over kwaliteit, bedoelen wij eigenlijk het deelaspect 'effectiveness', door Williamson gedefinieerd als 'the extent to which benefits, achievable under optimal conditions of care are actually achieved in clinic practice'.

Efficacy is gedefinieerd als 'the extent to which a health care intervention can be shown to be beneficial under optimal conditions of care'. Het gaat hierbij om de werkzaamheid van een (be)handeling onder laboratoriumomstandigheden, bij geselecteerde patiënten en vaak door zeer gespecialiseerde artsen uitgevoerd. Is hierover al onvoldoende bekend in de specialistische geneeskunde (Brook en Lohr, 1985), in de huisartsgeneeskunde is hiernaar helemaal weinig onderzoek gedaan (Knottnerus en De Vet, 1987).

Het begrip efficiency wordt gedefinieerd als 'the proportion of total cost (e.g. money, scarce resources and time) that can be related to actual benefits achieved'. Omdat we op zoek naar meetmethoden en instrumenten vooral geïnteresseerd zijn in de inhoudelijke kant van het geneeskundige werk, laten we de onderzoeken naar 'efficiency' hier buiten beschouwing. Overigens zijn wij van mening dat door het gebruik van een kwaliteitsbegrip, waarvan het begrip 'kosten' een belangrijk deel uitmaakt, het gevaar ontstaat dat verwarring van termen optreedt. Het gevaar dreigt namelijk dat een discussie die van financiële aard is, over de verdeling van middelen, gevoerd wordt onder het mom van 'kwaliteitsbewaking'.

Nog een begrenzing van dit literatuuronderzoek is, dat wij ons willen beperken tot de individuele hulpverleningssituatie. Begrippen die deelaspecten van kwaliteit weergeven, zoals equity en accessibility, worden gehanteerd om systemen van zorg te evalueren (Vuori, 1982) Deze begrippen vallen daarom buiten het hier gehanteerde kwaliteitsbegrip.

Zoals ook zal blijken bij de beschrijving van de verschillende gevonden meetmethoden, is er sprake van een continuum van situaties waarin het handelen van hulpverleners onderzocht kan worden, variërend van het werkelijke handelen in de dagelijkse praktijk tot het handelen in zeer specifieke situaties, test- of examensituaties. Aangezien wij op zoek zijn naar meetmethoden om het handelen van huisartsen in de dagelijkse praktijk te evalueren en tevens uit praktische overwegingen, is ook hier een grens getrokken, en wel zodanig dat de literatuur over onderwijssituaties en examens grotendeels buiten beschouwing is gebleven.

Nadat we het gebied waartoe het literatuuronderzoek zich zal beper-

ken op deze wijze hebben afgebakend, blijft er een aantal aspecten van kwaliteitsmeting over die ieder op zich aanleiding hebben gegeven tot een zeer groot aantal publikaties. Deze aspecten zijn: de arts-patiënt-relatie, gezondheidsstatus en satisfactie. Om aan deze literatuur voldoende recht te doen, zou aan ieder van deze onderwerpen een literatuuronderzoek gewijd moeten worden van minstens de omvang van het huidige onderzoek. Binnen het huidige tijdsbestek was dit niet haalbaar; daarom is volstaan met het in grote lijnen schetsen van de mogelijkheden en verwijzen naar andere overzichten.

Na de bespreking van al deze begrenzings zou de indruk kunnen ontstaan, dat er nauwelijks materiaal voor een literatuuronderzoek overblijft. Deze indruk is echter niet terecht, zoals zal blijken uit het vervolg van dit rapport. In het kort behandelt dit rapport de Nederlands- en Engelstalige literatuur over meetmethoden om kwaliteit in de zin van doeltreffendheid te meten van huisartsgeneeskundige zorg in de individuele hulpverleningssituatie.

Tot slot nog een opmerking over de weergave van de gevonden publikaties. Dit literatuuronderzoek is gericht op het vinden en evalueren van meetmethoden. Hierdoor worden de doelstellingen van de verschillende auteurs en de vermelde resultaten van de onderzoeken onderbelicht. Voor een meer gedetailleerde weergave van de volledige inhoud van de publikaties wordt de lezer verwezen naar de bijlage bij dit rapport en naar de bibliografie 'Kwaliteit van Huisartsgeneeskunde', die parallel aan dit rapport zal verschijnen.

De opbouw van dit rapport is als volgt: na een beschrijving van de gevolgde werkwijze bij het verzamelen van de literatuur in hoofdstuk 2, worden de gevonden meetmethoden beschreven aan de hand van een aantal verschillende dimensies die aan het meetproces onderscheiden kunnen worden (hoofdstuk 3 en 4). In hoofdstuk 5 wordt een beschouwing gegeven van de belangrijkste bevindingen en worden aanbevelingen gedaan voor verder onderzoek en ontwikkeling van meetmethoden. Een schematisch overzicht met de belangrijkste gegevens van alle opgenomen onderzoeken is toegevoegd als bijlage I. In de tekst zijn die onderzoeken, die in de bijlage zijn opgenomen, gemerkt met een *.

2. GEVOLGDE WERKWIJZE

Zoals ook al in de inleiding werd aangegeven, is de hoeveelheid literatuur over het meten van kwaliteit zeer omvangrijk. Van de literatuur die werd verzameld, is daarom een selectie gemaakt aan de hand van de in de inleiding beschreven begrenzingen. Bij de eerste verzameling van literatuur is uitgegaan van de 'Overzichtstudie onderzoek eerstelijns' (Sluijs e.a., 1985), de bibliografie 'Methoden en technieken binnen gezondheidszorgonderzoek' (Kuijlen, 1987) en de 'Registratie wetenschappelijk onderzoek eerstelijnsgezondheidszorg 1986' (Becht-Melai, 1987). Daarnaast is een literatuursearch gepleegd in een grote internationale database (Dimdi) aan de hand van trefwoorden, zoals 'quality', 'evaluation', 'measurement', 'criteria', gecombineerd met 'general practitioners' en 'primary care'. Tevens werden de laatste vijf jaargangen geraadpleegd van de volgende tijdschriften: Huisarts en Wetenschap, Medisch Contact, the Journal of the Royal College of General Practitioners, Medical Care, the American Journal of Public Health, Milbank Memorial Fund Quarterly, Social Science and Medicine, Quality Review Bulletin, the Journal of Medical Education. De literatuurverzameling is dus beperkt gebleven tot de Nederlands- en Engelstalige literatuur.

Aangezien deze methode niet specifiek genoeg bleek, werd uit de gevonden literatuur een selectie gemaakt op grond van de volgende overwegingen:

1. Is er sprake van een beschrijving van een onderzoek of onderzoeksmethode?
2. Gaat het om een wetenschappelijke methode om kwaliteit te meten?
3. Is de methode toepasbaar in de Nederlandse huisartsgeneeskunde?
4. Richt de methode zich op de individuele hulpverleningssituatie of is zij daarvoor toepasbaar?
5. Richt de methode zich op het meten van kwaliteit onder normale omstandigheden?

De achtergronden en implicaties van deze overwegingen zijn de volgende:

Ad. 1.:

Dit literatuuronderzoek richt zich op het vinden van meetmethoden; dit houdt in dat minder aandacht is besteed aan editorials en discussiestukken.

Ad. 2.:

Door dit selectie criterium wordt een grote hoeveelheid literatuur over kwaliteitsbevordering en kwaliteitsbewaking uitgesloten; deze is meer gericht op het ontwikkelingsaspect van een bepaald programma (hoe verbeter je de kwaliteit van het handelen van hulpverleners?) dan op het evaluatie-aspect. Daarnaast heeft deze overweging de bedoeling een onderscheid te maken tussen inventariserend onderzoek, bijvoorbeeld naar taakopvatting en werkwijze van huisartsen, en onderzoek waarin werkelijk sprake is van het meten van kwaliteit van huisartsgeneeskundig handelen. Dit onderscheid is overigens vaak lang niet zo duidelijk als het lijkt.

Ad. 3.:

De doelstelling van dit literatuuronderzoek richt zich op meetmethoden die toepasbaar zijn in de Nederlandse huisartsgeneeskunde; in de inleiding werd al duidelijk dat met name in de Amerikaanse literatuur de toepasbaarheid in de Nederlandse huisartsgeneeskunde niet altijd op het eerste gezicht duidelijk is.

Ad. 4.:

De 'individuele hulpverlening' wordt hier beschouwd als tegenstelling van 'systeem van zorg'; voor de evaluatie van systemen spelen andere kwaliteitsaspecten en criteria een rol, zoals bijvoorbeeld 'kosten'.

Als gevolg van onze keuze dit literatuuronderzoek te richten op de individuele hulpverleningssituatie en niet op systemen van zorg, blijft de literatuur over het meten van kwaliteit op het niveau van structure hier buiten beschouwing.

Ad. 5.:

Zoals in de inleiding werd aangegeven, worden laboratorium- en examensituaties uitgesloten.

Het zal duidelijk zijn dat de selectie die op deze wijze is gemaakt, gebaseerd is op een subjectief oordeel. Het is bovendien uitgesloten, dat op deze wijze een compleet overzicht van alle publikaties over het meten van kwaliteit in de huisartsgeneeskunde ontstaat. Ondanks deze tekortkomingen is er naar gestreefd in elk geval een zo volledig mogelijk beeld te krijgen van de belangrijkste bestaande meetmethoden, teneinde aanbevelingen te kunnen doen omtrent meer en minder wenselijke aspecten van verschillende methoden.

3. KWALITEITSMETING OP VERSCHILLENDE NIVEAU'S

Zoals in de inleiding werd aangegeven, kan men volgens de indeling van Donabedian kwaliteit meten op drie niveau's, te weten structure, process en outcome. De achterliggende gedachte hierbij is, dat het resultaat (outcome), het uiteindelijke doel van de hulpverlening, het gevolg is van wat er zich in het hulpverleningsproces afspeelt. Dat hulpverleningsproces is op zijn beurt weer afhankelijk van de aanwezige voorwaarden, structure.

De kwaliteit van de individuele hulpverleningssituatie kan worden gemeten op process- of outcomeniveau en wordt soms ook in één onderzoek gemeten op beide niveau's. Als men de kwaliteit van het handelen van huisartsen wil meten, zou men eigenlijk een beeld moeten hebben van het totale handelen van een huisarts. Een groot probleem hierbij is de veelomvattendheid van 'het handelen van de huisarts'. De huisarts krijgt te maken met een groot aantal verschillende hulpvragen en verricht een groot aantal zeer uiteenlopende handelingen. Aangezien het in praktische zin onmogelijk is een meetinstrument te ontwikkelen waarin alle facetten van het handelen van de huisarts evenredig aan bod komen, beperken veel onderzoekers zich tot bepaalde onderdelen ervan. Soms richt de aandacht zich op bepaalde handelingen, soms op bepaalde soorten hulpvragen (vaak klachten of diagnoses).

Sommige auteurs gaan ervan uit dat dit soort metingen aan een beperkt deel van het handelen van de huisarts een valide beeld geeft van het totale handelen van de huisarts; anderen houden hun conclusies beperkt tot dat onderdeel waarop zij zich met hun meetinstrument gericht hebben. Deze verschillende mogelijkheden worden in dit hoofdstuk beschreven aan de hand van de gevonden literatuur. Eerst worden onderzoeken besproken waarin kwaliteit wordt gemeten op processniveau (paragraaf 3.1), vervolgens outcomeniveau (paragraaf 3.2) en tot slot onderzoeken waarin op beide niveau's wordt gemeten (paragraaf 3.3).

3.1. Kwaliteitsmeting op processniveau

Bij meting op processniveau is de aandacht direct gericht op 'het handelen van de huisarts'. Hierin kan men het handelen bij verschillende soorten hulpvragen onderscheiden, hetgeen in onderzoek vaak

betekent dat men zich beperkt tot bepaalde klachten of bepaalde aandoeningen. Een ander onderscheid dat gemaakt kan worden, is tussen verschillende soorten handelingen. Analoog aan de indeling die door de Nationale Raad wordt voorgesteld (zie inleiding) kan men het handelen van de huisarts indelen in: medisch-technisch, psychosociaal en organisatorisch. Medisch-technisch handelen is gebaseerd op het vakkundige, vooral cognitieve deel van het werk; met psychosociaal handelen doelen wij op het hanteren van de arts-patiënt-relatie en organisatorisch handelen heeft te maken met zaken als regeling van continue bereikbaarheid, wachttijden op het spreekuur, etcetera. Binnen deze indeling kan men een nog meer gedetailleerde onderverdeling maken, zodat het medisch-technisch handelen bijvoorbeeld als volgt kan worden uitgesplitst: anamnese, lichamelijk onderzoek, aanvullend laboratorium- en röntgenonderzoek, diagnose-stelling, verwijzen, therapie, informatie verstrekken, evaluatie van de therapie. Voor psychosociaal en organisatorisch handelen kan men soortgelijke indelingen maken. Deze indelingen zijn niet absoluut; bijvoorbeeld aan de handeling 'verwijzen' kan men een medisch-technisch aspect onderscheiden (wordt de juiste patiënt op het juiste moment verwezen?), maar ook een organisatorisch aspect (wordt de verwijzkaart leesbaar ingevuld en wordt voldoende informatie aan de specialist verstrekt?). Het is niet de bedoeling hier deze indeling tot in details uit te werken. De indeling wordt gebruikt als leidraad bij de bespreking van de literatuur in deze paragraaf.

Door gebruik te maken van de methode waarbij een **impliciet** oordeel gegeven wordt over de kwaliteit van het handelen van een huisarts, kan men het probleem, dat er geen meetinstrument te maken is dat het volledige handelen van de huisarts bestrijkt, omzeilen. Het zorgproces wordt dan gedurende een bepaalde tijd of tijdens een aantal consulten direct beoordeeld door één of meer collega's of deskundigen. Deze geven dan een - al dan niet gespecificeerd - impliciet oordeel over de kwaliteit van wat zij hebben gezien. In zijn meest simpele vorm (men kan zich zelfs afvragen of hier sprake is van kwaliteitsmeting of alleen van waarneming) wordt deze methode beschreven door Plaisier. Hij woonde een aantal spreekuren bij van collega's en gaf een impliciet oordeel over verschillende aspecten van wat hij gezien had (Plaisier, 1987*). In het rapport 'What sort of doctor?' van de Royal College of General Practitioners wordt een beschrijving gegeven van een proefonderzoek volgens deze methode van impliciete beoordeling (Royal College of General Practitioners, 1985). Voor dit proefonderzoek was een aantal zaken vastgelegd, maar de beoordeling

bleef impliciet. Van tevoren werd vastgesteld welke aspecten beoordeeld zouden worden, verdeeld over vier categorieën (professionele waarden, toegankelijkheid, klinische competentie en communicatieve vaardigheden) en op welke wijze informatie verzameld zou worden (vragenlijst aan de huisarts, directe observatie, gesprekken met de praktijkmedewerkers, beoordeling van een steekproef van de patiëntenkaarten, video-opnamen van consulten en een interview met de huisarts). Naast process-aspecten werden ook aspecten op het niveau van 'structure' beoordeeld. De bezoekende collega-huisartsen stelden na afloop van de beoordeling een gedetailleerd verslag op over hun bevindingen, dat aan de beoordeelde huisarts werd toegestuurd. Door de deelnemende huisartsen werd de validiteit en betrouwbaarheid van de methode weliswaar positief beoordeeld, maar naar onze mening kunnen hierbij wel enige vraagtekens worden geplaatst. Een nadere beschouwing van impliciete kwaliteitsbeoordeling volgt in paragraaf 4.2.

Een bepaald onderdeel van het medisch-technisch handelen van de huisarts kan ook centraal staan. Zo wordt in een tweetal onderzoeken de **werkwijze en hypothesevorming** van huisartsen en internisten vergeleken (Gerritsma en Smal, 1982*; Simpson e.a., 1987*). Hiervoor wordt gebruik gemaakt van zogenaamde 'papierpatiënten'. De keuzes die de arts maakt tijdens het voor een bepaalde casus nagebootste hulpverleningsproces worden nauwkeurig vastgelegd. Door deze beide onderzoeken wordt het verschil in denkwijze en probleemoplossend gedrag tussen huisartsen en internisten zichtbaar gemaakt, zonder dat een oordeel wordt gegeven over welke van de twee beter is. Internisten genereren over het algemeen meer hypothesen wanneer zij met een diagnostisch probleem geconfronteerd worden en gaan vervolgens proberen zoveel mogelijk van die hypothesen te weerleggen; zij trachten de kans op 'fouten van omissie' zo klein mogelijk te maken. Huisartsen daarentegen gaan in beginsel vooral uit van enkele, meest waarschijnlijke hypothesen, waarna al snel één waarschijnlijkheidsdiagnose geformuleerd wordt. Daar dit verschil in werkwijze zeker ook het verschil tussen de 'setting' van de huisartsenpraktijk en de (poli)klinische populatie weerspiegelt, kan niet worden beargumenteerd dat het gevonden verschil een kwaliteitsverschil inhoudt. Instrumenten, zoals ontwikkeld door Gerritsma en Smal, zouden echter ook gebruikt kunnen worden om huisartsen onderling te vergelijken, mits criteria kunnen worden vastgesteld die specifiek voor de huisartsgeneeskunde geldig zijn. Ook door Kuyvenhoven e.a. werd gebruik gemaakt van de methode van papierpatiënten om specifieke onderde-

len van het handelen van de huisarts te onderzoeken. Het ging hierbij om: aandacht voor relevante somatische aspecten, oorzaken van vage klachten en de kans op onnodige iatrogene schade (Kuyvenhoven e.a., 1984a; Jacobs e.a., 1984, Kuyvenhoven e.a., 1986*). Meer hierover in paragraaf 4.1.

Een ander aspect van het medisch-technisch handelen van de huisarts dat als onderwerp van onderzoek heeft gediend, is het **verwijzen**. Door Grace en Armstrong werd de kwaliteit van het verwijsgedrag gemeten op een zeer basaal niveau, namelijk door telkens de huisarts, patiënt en specialist die betrokken waren bij een verwijzingsepisode, eenzelfde vragenlijst voor te leggen omtrent de juistheid van het handelen van de huisarts in dat specifieke geval (Grace en Armstrong, 1987*). De betrouwbaarheid van een dergelijke methode is moeilijk na te gaan omdat het steeds een mening betreft van individuele personen omtrent individuele episodes. Bovendien kan men zich afvragen of de mening van één van de drie betrokkenen danwel van alle drie tezamen een valide kwaliteitsmaat is. Toch krijgt men op deze wijze wel een indruk van (het gebrek aan) overeenstemming over criteria van juist verwijsgedrag tussen de verschillende betrokkenen. In slechts 21,1% van de gevallen waren alle drie de betrokkenen het erover eens dat de huisarts niet meer had kunnen doen, alvorens te verwijzen.

In een drietal onderzoeken uit Groot-Brittannië worden aan de hand van gegevens uit patiëntenstatussen specifieke aspecten van het handelen van huisartsen beoordeeld. Uit de beschrijvingen wordt echter niet duidelijk of, en zo ja, welke criteria gehanteerd zijn. De beschrijvingen maken meer de indruk van inventarisaties waarbij achteraf min of meer impliciet de kwaliteit beoordeeld wordt. Het gaat hierbij respectievelijk om **voorschrijven** bij de ziekte van Parkinson (Wilson & Murray, 1985*), **screening** op cervixcarcinoom (Wilson, 1987*) en **preventie** in de huisartsenpraktijk in het algemeen (Fleming en Lawrence, 1983*). In dit laatste onderzoek was er sprake van een vergelijking die werd gemaakt met een eerder uitgevoerde 'inventarisatie' in dezelfde 29 huisartsenpraktijken 2 1/2 jaar eerder. Geconcludeerd werd dat het preventieve handelen van de huisartsen was verbeterd.

Over **voorschrijven** gaan ook de Nederlandse onderzoeken van Haaijer-Ruskamp en van Morkink (Haaijer-Ruskamp e.a., 1983*; Haaijer-Ruskamp, 1984; Morkink, 1986). Deze worden verder behandeld in paragraaf 4.1.

Door Imants werd in Nederland een onderzoek uitgevoerd naar de prak-

tische uitvoering van **bloeddrukmeting** door huisartsen (Imants, 1979*). Als onderzoeksmateriaal werden video-banden gebruikt die al eerder ten behoeve van een ander onderzoek waren opgenomen. Bij de materiaalverzameling waren de huisartsen dus niet op de hoogte van het feit dat ze op dit specifieke onderdeel beoordeeld zouden worden. Aan de hand van expliciete normen (richtlijnen van het Nederlands Huisartsen Instituut) werd het proces van bloeddruk meten bij patiënten door huisartsen beoordeeld. Alhoewel het hier slechts gaat om een minimaal onderdeel van het huisartsgeneeskundig handelen, is naar onze mening een dergelijke beoordeling van hoe een bepaalde handeling wordt uitgevoerd een meer valide benadering van het begrip 'kwaliteit van het handelen' dan de registratie van de uitkomst van een bloeddrukmeting in de status.

In een aantal onderzoeken wordt de aandacht gericht op bepaalde aspecten van het handelen van de huisarts die meer passen in de categorie **organisatorisch handelen**. Fletcher e.a. trachtten in een onderzoek naar **verwijzen** de continuïteit en coördinatie van het verwijzen te meten. Continuïteit werd geoperationaliseerd als het percentage van alle consulten dat werd gedaan door de 'eigen' arts. Coördinatie bestond uit drie componenten, die allen een geschreven bewijs reflecteerden dat de eigen arts ofwel op de hoogte was van het feit dat een andere hulpverlener werd geconsulteerd, ofwel daarbij betrokken was (Fletcher e.a., 1984*).

In dit onderzoek gaat het echter om verwijzingen binnen een instelling; het is dus niet zonder meer toepasbaar op de Nederlandse situatie. Toch worden de begrippen continuïteit en coördinatie ook in de Nederlandse eerstelijns steeds belangrijker, naarmate het aantal groepspraktijken en gezondheidscentra groeit en de samenwerking tussen de verschillende disciplines (en tussen de echelons) toeneemt. Holmes e.a. deden een poging begrippen, die zij beschouwen als kenmerkend voor 'family practice', te operationaliseren en te meten. Deze begrippen zijn: continuïteit, coördinatie, veelomvattendheid en gerichtheid op het gezin. Vergeleken werden huisartsen die wel met huisartsen die niet een speciaal op 'family practice' gerichte opleiding hadden genoten. Om de vier begrippen te meten, werden expliciete criteria vastgesteld, zonder dat er een waarde-oordeel aan was verbonden. Continuïteit werd geoperationaliseerd als het percentage bezoeken aan de eigen arts, samen met het percentage herhaalbezoeken op verzoek van de arts waaraan patiënten voldeden; coördinatie werd geoperationaliseerd als het percentage verwijzingen die resulteerden in informatie terug naar de aanvrager; voor veelomvattendheid gold

het percentage bezoeken waarbij informatie werd vastgelegd over de gezondheidsstatus van de patiënt buiten de hoofdklacht, samen met het gemiddelde aantal notities over de gezondheid in een steekproef van statussen; gezinsgerichtheid tenslotte werd bepaald aan het gemiddelde aantal notities over de familie-anamnese en het percentage van alle gezinsleden van een bepaald aantal patiënten dat de huisarts bezocht in een bepaalde periode (Holmes e.a., 1978*).

Binnen het evaluatie-onderzoek van het experiment Diagnostisch Centrum Maastricht werd een onderzoek gedaan naar de doelmatigheid van het **aanvragen van diagnostiek** (Brouwer e.a., 1983*). Na bestudering van de aanvraagformulieren voor diagnostiek, zoals deze door huisartsen waren ingestuurd, werden in overleg met de betrokkenen werkafspraken gemaakt over hoe aanvragen eruit zouden moeten zien en voor welke indicaties aanvragen gedaan zouden worden. Vervolgens werden de nieuwe aanvraagformulieren beoordeeld aan de hand van deze werkafspraken. Hoewel hier slechts een klein onderdeelje van het handelen van huisartsen betreft, is er duidelijk sprake van kwaliteitsmeting in de zin van 'de werkelijkheid toetsen aan vooraf vastgestelde criteria'. Een bezwaar dat men tegen dit onderzoek zou kunnen aanvoeren, is dat de beoordeelde huisartsen op de hoogte waren van het feit dat ze beoordeeld werden, waarop en volgens welke normen, en dat er dus sprake zou kunnen zijn van een 'performance gap' (in dit concrete geval verbeterden de huisartsen hun gedrag trouwens nauwelijks, zodat dit argument hier niet op lijkt te gaan).

Nadat wij in het voorgaande een aantal onderzoeken besproken hebben waarin een bepaald onderdeel van het handelen van huisartsen werd beoordeeld, willen wij nu onderzoeken beschouwen die zich richten op **bepaalde aandoeningen**.

Door Kessner e.a. is de zogenaamde **tracertechniek** ontwikkeld. Hierbij wordt aangenomen dat routinematig verleende zorg voor veel voorkomende aandoeningen, zogenaamde tracers, beschouwd kan worden als een maat voor de gehele zorgverlening van een hulpverlener of instelling. Voor het selecteren van traceraandoeningen wordt een aantal expliciete criteria gegeven; de beoordeling van de hulpverlening voor deze aandoeningen vindt plaats op basis van gegevens uit statussen aan de hand van (andere) expliciete criteria (Kessner e.a., 1973; Kessner, 1978). Hoewel in buitenlands onderzoek succesvolle toepassing van de techniek is beschreven (Sibley e.a., 1975*; Novick e.a., 1976*; Smith, 1977*), blijkt uit een beschrijving van de problemen bij een poging de methode toe te passen door De Haen en Wijkkel toepassing in de Nederlandse eerstelijns niet eenvoudig. Afge-

zien van het feit dat men de geldigheid van het uitgangspunt van de tracermethode, dat routinematig verleende zorg een indicator is van de totale zorgverlening, in twijfel kan trekken, blijken er ook een aantal praktische problemen te zijn. De informatie op de 'groene kaart' van huisartsen blijkt onvoldoende om de methode goed te kunnen toepassen. Daarnaast is het niet eenvoudig een traceraandoening te vinden die voldoet aan alle door Kessner e.a. geformuleerde criteria. Ook kan men zich afvragen of het juist is in de huisartsgeneeskunde uit te gaan van diagnoses; patiënten komen immers met klachten en de criteria voor het stellen (en noteren) van bepaalde diagnoses kunnen nogal variëren tussen huisartsen. Tot slot kan nog worden opgemerkt dat de tracermethode alleen de medisch-technische kant van de hulpverlening evalueert; de psychosociale kant wordt aanzienlijk onderbelicht (De Haen en Wijkel, 1982).

Er is nog een aantal andere onderzoeken gevonden die de zorg voor bepaalde aandoeningen (in aantal variërend van één tot 180) evalueren met behulp van informatie uit de statussen en aan de hand van expliciete criteria, echter zonder dat men pretendeert de totale zorgverlening te kunnen beoordelen (Bennett e.a., 1985*; Borgiel e.a., 1985*; Palmer e.a., 1985; Rudnick e.a., 1983*; Sheps en Robertson, 1984*). In al deze onderzoeken wordt weer voornamelijk de medisch-technische kant van de hulpverlening beoordeeld.

Ook onderzoeken met behulp van simulatiepatiënten richten zich op één of slechts enkele aandoeningen, bijvoorbeeld Rethans en Van Boven, 1987*, 1988; Norman e.a., 1985*. Meer hierover in paragraaf 4.1.

Voor het beoordelen van alleen de **psychosociale kant** van de hulpverlening van huisartsen is een zeer groot aantal instrumenten ontwikkeld. Deze variëren van meetinstrumenten waarin zeer gedetailleerd het gedrag van arts en patiënt wordt vastgelegd (bijvoorbeeld Brannigan en Humphries, geciteerd in Fassnacht, 1982) tot globale beoordelingsschalen over bijvoorbeeld de houding van de arts ten opzichte van de patiënt (bijvoorbeeld Kraan en Crijnen, 1987).

Zoals ook al in de inleiding werd aangegeven, voert een uitgebreide beschrijving van al deze mogelijke instrumenten te ver; daarom wordt verwezen naar overzichten, zoals die van DiMatteo en DiNicola (1981) en Pendleton en Hasler (1983) en de NIVEL-bibliografie over arts-patiëntcommunicatie (De Leeuw en Vos, 1987). Hoewel er wel pogingen gedaan worden om de uitkomsten van dergelijk onderzoek naar arts-patiëntcommunicatie te valideren aan de hand van uitkomsten, zoals compliance en satisfactie, is het opvallend dat zelden een verband

wordt gelegd met meetmethoden van kwaliteit van medisch-technisch handelen.

Slechts enkele onderzoeken hebben wij gevonden waarin gepoogd wordt een **verband** te leggen tussen meetinstrumenten van **psychosociaal handelen** en van **medisch-technisch handelen**.

Een poging om een verband te leggen tussen de communicatieve aspecten van arts-patiënt-interacties en huisartsgeneeskundige kwaliteit is gedaan door Thomassen. Deze paste op 315 opnames van spreekuurcontacten zowel een beoordelingsinstrument voor communicatie toe als één voor kwaliteit, terwijl over het laatste begrip bovendien bij een steekproef van 65 opnames een impliciet oordeel werd verkregen van huisartsen-deskundigen ter validering (Thomassen, 1986*). Geconcludeerd werd dat er een verband bestaat tussen communicatieve aspecten en huisartsgeneeskundige kwaliteit van het hulpverleningscontact.

Ook door het Nijmeegs Universiteit Huisartsen Instituut is onderzoek gedaan naar het handelen van huisartsen, waarbij zowel medisch-technische als psychosociale aspecten werden beoordeeld. Uitgaande van materiaal dat verzameld is in het kader van het project 'Preventie van Somatische Fixatie', is een aantal onderzoeken gepubliceerd die, hoewel de term 'kwaliteit' door de auteurs zelf relatief weinig gebezigd wordt, zeker een belangrijke plaats innemen in het onderzoek naar kwaliteit in de Nederlandse huisartsgeneeskunde (Grol e.a., 1985a; Grol e.a., 1985b; Grol e.a., 1986a*; Grol e.a., 1986b*; Grol, 1987*; Grol e.a., 1988*; Mokkink, 1986; Mokkink e.a., 1986*; Smits e.a., 1985*; Tielens e.a., 1986*). Van de consulten van een aantal huisartsen (overigens geen aselechte groep), waarover informatie werd verzameld door directe observatie en audio-opnamen, werd de medisch-technische kant beoordeeld aan de hand van protocollen die waren ontwikkeld voor 24 nieuwe klachten (dus niet: diagnoses). Meer over deze protocollen in paragraaf 4.2. Voor de beoordeling van het 'niet-medisch handelen' werd gebruik gemaakt van impliciete beoordeling van de aspecten 'gericht en systematisch werken', 'hanteren van de huisarts-patiënt-relatie' en 'psychosociaal handelen'. Zoals te verwachten, was de interbeoordelaarbetrouwbaarheid van de meting van het medisch-technisch handelen aan de hand van expliciete criteria hoger ($\kappa = 0,71-0,75$) dan van het niet-medisch handelen aan de hand van impliciete criteria (Kendall's tau = $0,30-0,60$). Met behulp van deze meetinstrumenten, gecombineerd met andere gegevens, werd onder andere de samenhang van de kwaliteit van het handelen met attitudes en werksatisfactie van huisartsen onderzocht (Grol e.a.,

1985a; Grol e.a., 1985b), evenals het effect van een programma van onderlinge toetsing (Grol e.a., 1986a*; 1986b*; 1988*; Grol, 1987*). Ook werd hiermee onderzoek gedaan naar het verband tussen patiëntgericht werken en adequate diagnostiek en therapie en naar het verschil in kwaliteit van het handelen bij mannelijke en vrouwelijke patiënten (Smits e.a., 1985*; Tielens e.a., 1986*). Door Mokkink werd op grond van uitkomsten van deze twee instrumenten, gecombineerd met ziekenfondscijfers over voorschrijven en verwijzen, een indeling gemaakt in verschillende werkstijlen van huisartsen, gebaseerd op onderlinge vergelijking (percentielen) (Mokkink e.a., 1986*; Mokkink, 1986). Alhoewel hier consequent gesproken wordt over werkstijlen en niet over kwaliteit, kan men zich nauwelijks onttrekken aan de indruk dat het hier in werkelijkheid om kwaliteitsverschillen gaat.

De hier besproken meetinstrumenten zijn uitgebreid uitgewerkt en getest en maken een valide indruk. Daarom is het ons inziens hoog tijd dat ze ook in een andere, zo mogelijk aselekt gekozen steekproef van huisartsen, worden toegepast. Al deze publikaties zijn namelijk gebaseerd op onderzoeksresultaten bij dezelfde groep huisartsen. Voor verdere ontwikkeling en validering van de instrumenten zou het bevorderlijk zijn als ze op grotere schaal zouden worden uitgetest.

Samenvattend kan men over het onderzoek naar kwaliteit in de huisartsgeneeskunde op processniveau het volgende opmerken: er is een behoorlijk aantal verschillende methoden ontwikkeld en toegepast om kwaliteit van huisartsgeneeskundige zorg te meten op processniveau. Wat betreft het onderdeel van het handelen van huisartsen dat wordt geëvalueerd, blijkt er een kloof te bestaan tussen enerzijds meetinstrumenten om de psychosociale kant van het handelen te beoordelen (waarvan een gedetailleerde beschrijving buiten de omvang van dit literatuuroverzicht valt) en anderzijds meetinstrumenten naar de medisch-technische kant. Slechts in enkele onderzoeken wordt een poging beschreven deze twee aspecten samen en in hun onderlinge samenhang te meten. Een veelbelovende methode hiervoor is ontwikkeld door het Nijmeegs Universitair Huisartsen Instituut. Daar echter alle publikaties over deze methode tot nu toe betrekking hebben op de meetresultaten van één, niet-aselecte groep huisartsen, wordt ervoor gepleit de methode ook in andere, liefst aselekt samengestelde groepen huisartsen uit te testen.

3.2. Kwaliteitsmeting op outcomeniveau

Met de 'outcome' van hulpverlening bedoelt Donabedian het resultaat van de hulpverlening; datgene waar het uiteindelijk om te doen is. Outcome kan men meten als positief resultaat, of als (het vóórkomen van) ongewenste resultaten of complicaties, met als uiterste grens mortaliteit. Vergelijkbaar met het processniveau kan men bij outcome een medisch-technische en een psychosociale kant onderscheiden. Deze verschillende mogelijkheden worden geïllustreerd aan de hand van de gevonden literatuur.

Een **medisch-technische outcome** die relatief gemakkelijk meetbaar is, is **mortaliteit**. In de huisartsgeneeskunde gaat het echter in het merendeel van de gevallen om weinig levensbedreigende aandoeningen, zodat het meten van mortaliteit meestal niet relevant is. Wel relevant is het indien men zich beperkt tot een groep patiënten met een levensbedreigende aandoening, zoals werd gedaan door Hill e.a., die zich richtten op patiënten verdacht van myocardinfarct (Hill e.a., 1978*). Een groep van patiënten verdacht van myocardinfarct en uitgebreid gescreend op mogelijke complicaties (een uitsluitingscriterium voor het onderzoek) werd aselekt verdeeld in twee groepen. De ene groep werd door de huisarts begeleid en de andere groep werd opgenomen, waarna na zes weken de outcome werd vastgelegd, onder andere in termen van mortaliteit. De aselechte toewijzing aan één van beide groepen vergroot de betrouwbaarheid van de resultaten uit dit onderzoek aanzienlijk.

Een ander onderzoek, waarin gekeken werd naar '**ongewenste resultaten**', is dat van Kaptein e.a. naar de kwaliteit van hulpverlening door huisartsen bij asthma (Kaptein e.a., 1986; 1987*). In dit onderzoek werd de outcome bepaald door middel van interviews met patiënten. Hun werd gevraagd naar het optreden van klachten in de maand voorafgaand aan het onderzoek en naar de voorgeschreven medicatie. Op grond van verschillende categorieën van combinaties van klachten en medicatie waren van tevoren expliciete criteria opgesteld voor **onderbehandeling**. Men kan zich hierbij afvragen in hoeverre van de patiënt verkregen informatie over voorschriften voor medicatie juist is; het is immers mogelijk dat de patiënt de voorschriften verkeerd heeft begrepen, of zich er niet aan houdt.

Gezondheidsstatus is een 'medisch-technische outcome' waarover een grote hoeveelheid literatuur bestaat. Meetinstrumenten voor gezondheidsstatus kunnen vaak goed gebruikt worden voor onderzoek naar kwaliteit, maar zijn niet specifiek daarvoor bedoeld. De literatuur over deze instrumenten is dan ook vaak niet op deze toepassing (van

kwaliteitsonderzoek) gericht. Daarom wordt hier volstaan met een zeer summier overzicht van de verschillende mogelijkheden.

Bij vragenlijsten voor 'patiënten' is het vaak niet mogelijk lichamelijke en psychosociale factoren strikt te scheiden. Door Gonella en Goran is een puur lichamelijke schaal voorgesteld, gebaseerd op medische diagnostiek, het zogenaamde 'staging concept'. Aangezien het hierbij gaat om soms zeer verfijnde diagnoses, is deze schaal voor de huisartspraktijk minder bruikbaar (Gonella en Goran, 1975). Door Meyboom-de Jong e.a. is in Nederland een voor de eerstelijns aangepaste versie van the Rand Health Insurance Study Battery toegepast (Meyboom-de Jong e.a., 1986a, 1986b). Dit meetinstrument gaat uit van wat de patiënt ervaart. Andere meetinstrumenten voor '**functionele status**' zijn bijvoorbeeld the Sickness Impact Profile en the General Health Questionnaire (Hall e.a., 1987).

Ook voor het meten van **psychische gezondheid** of welbevinden, zoals door de patiënt wordt ervaren, is een groot aantal meetinstrumenten ontwikkeld. Door Hawton e.a. werden bij een onderzoek naar de eventuele verschillen in kwaliteit bij de begeleiding van patiënten die een zelfmoordpoging hadden gedaan door huisartsen en bij een polikliniek, naast beoordeling door de huisarts en door een onderzoeker, vijf verschillende meetinstrumenten gehanteerd, die werden voorgesteld aan de patiënten (Suicidal Intent Scale, Risk of Repetition Scale, Social Adjustment Scale, Beck Depression Inventory en General Health Questionnaire) (Hawton e.a., 1987*).

Naast psychische gezondheid of welbevinden is **satisfactie** een maat voor 'psychosociale outcome'. Een combinatie van medisch-technische outcome-meting en psychosociale outcome-meting vonden wij in een publikatie van Crebolder en Van Eijndhoven. Dit onderzoek richtte zich op de zorg van de huisarts voor type II diabeten (Crebolder en Van Eijndhoven, 1986*). Naast het meten van medisch-technische outcome, de instelling van de diabetes (bepaald aan het nuchter bloedsuikergehalte, de quetelet-index en het HbA1c-percentages) werd ook een oordeel van de patiënten gevraagd over de geleverde zorg (satisfactie).

Zoals al in de inleiding werd aangegeven, zullen wij de literatuur over **satisfactie** slechts kort aanduiden.

Patiëntensatisfactie is een moeilijk meetbaar begrip. Opvallend is - vooral bij weinig genuanceerd gestelde vragen - het lage percentage 'patiënten', dat aangeeft niet tevreden te zijn (Donabedian, 1985).

Door vragen te stellen over tevredenheid met meer verschillende aspecten van de zorg, kan men een beter beeld krijgen van het oordeel

van patiënten; dergelijke aspecten zijn bijvoorbeeld de volgende: competentie van de arts, vriendelijkheid, persoonlijke zorg, sfeer, wachttijd, kosten (Dutton e.a., 1985). Behalve literatuur over de ontwikkeling en evaluatie van meetinstrumenten naar satisfactie, bijvoorbeeld Hulka e.a., 1970; Ware e.a., 1978 (aangehaald in Dutton e.a., 1985); Roberts en Tugwell, 1987; Bollam e.a., 1988, is er ook onderzoek gedaan naar factoren die satisfactie beïnvloeden, bijvoorbeeld structuurvariabelen zoals 'disability' (Patrick, 1983). Ook is satisfactie gehanteerd als onafhankelijke variabele om continuïteit van de hulpverlening te voorspellen (Marquis e.a., 1983). Voor een overzicht van de literatuur over patiëntensatisfactie verwijzen wij naar een tweetal recente artikelen van Hall en Dornan (Hall en Dornan, 1988a; 1988b).

In deze paragraaf hebben we gezien dat outcome gemeten kan worden op 'zuiver medisch' niveau, functioneel niveau (gezondheidsstatus) en op het niveau van patiëntensatisfactie. Bij de laatste twee metingen gaat het om het oordeel van de patiënt, in het eerste geval wordt de outcome bepaald door een arts, op medische gronden. Het maakt daarbij voor de betrouwbaarheid nogal wat uit hoe men tot dit oordeel komt, bijvoorbeeld of men op de hoogte is van wat er met de patiënt gebeurd is voor de outcomemeting; de beoordelaar kan hierdoor beïnvloed worden. Elke outcome die men meet (medisch, functioneel of satisfactie) wordt door verschillende factoren bepaald, die niet alleen van de hulpverlener afhankelijk zijn. Naast de invloed van de hulpverlener kunnen de volgende factoren een rol spelen: de toestand van de patiënt voor het begin van de hulpverlening (bijvoorbeeld gezondheidstoestand, complicerende factoren, verwachtingen), de invloed van de patiënt op het hulpverleningsproces (bijvoorbeeld mondigheid, begrip, medewerking, compliantie) en min of meer externe factoren waarop de hulpverlener, en vaak ook de patiënt, geen invloed heeft (bijvoorbeeld hygiënische omstandigheden, sociale omstandigheden, andere lichamelijke aandoeningen). Al deze mogelijke beïnvloedende factoren dienen in ogenschouw te worden genomen bij het beoordelen van de validiteit van outcomematen als indicatoren van de kwaliteit van de hulpverlening. Zo is ook bijvoorbeeld het tijdstip waarop de outcome wordt bepaald van belang: meet men te vroeg, dan heeft bijvoorbeeld een therapie nog geen effect; meet men te laat, dan is de invloed van allerlei externe factoren zo groot, dat die van de hulpverlening in het niet valt.

3.3. Meting van kwaliteit op zowel process- als outcomeniveau

Net zoals er onderzoeken gepubliceerd zijn waarin kwaliteit alleen op processniveau of alleen op outcomeniveau wordt gemeten, zijn er onderzoeken waarin op beide niveau's wordt gemeten zonder dat er verband wordt gelegd tussen de niveau's. In één onderzoek dat wij vonden, is zelfs op drie niveau's 'gemeten', namelijk een onderzoek van Kekki, waarin de kwaliteit van zorg in een groepspraktijk wordt beoordeeld (Kekki, 1987*). Uittreksels van statussen worden beoordeeld op verschillende aspecten bij verschillende handelingen en aandoeningen; wat de criteria waren, wordt echter uit de beschrijving niet duidelijk, evenmin als het precieze vooropgezette doel van deze evaluatie. De auteur beveelt diepergaand onderzoek aan naar de door hem gevonden tekortkomingen.

De beschrijving van een onderzoek van Payne e.a. naar het verschil tussen wel en niet gecertificeerde en oudere versus jongere artsen is heel wat duidelijker (Payne e.a., 1984*). Voor een aantal vastgestelde aandoeningen zijn van tevoren expliciete optimumcriteria op process- en outcomeniveau vastgelegd, waaraan de kwaliteit van de zorg wordt afgemeten.

Een tweetal onderzoeken waarin de kwaliteit van zorg voor geïsoleerde aandoeningen wordt geëvalueerd in steeds één huisartsenpraktijk, zijn die van Edwards e.a. naar de hulpverlening bij buikklachten en van Taylor naar het (medische)beleid bij epilepsie (Edwards e.a., 1985*; Taylor, 1987*). In geen van beide gevallen zijn duidelijke criteria vastgelegd van wat wel/niet goede zorg is; de resultaten worden in beschrijvende zin gegeven en daarna min of meer impliciet op 'kwaliteit' beoordeeld.

Een dergelijke procedure (beschrijvende resultaten en vervolgens een impliciet oordeel over de kwaliteit van zorg) is te vinden in een drietal onderzoeken naar het hypertensiebeleid in huisartsenpraktijken (Kurji en Haines, 1984*; Michael, 1984*; Van Ree e.a., 1985*). In een drietal onderzoeken naar de kwaliteit van huisartsgeneeskundig handelen bij diabetespatiënten werden evenmin expliciete criteria vastgelegd, maar werden vergelijkingen gemaakt: met resultaten van eerder onderzoek in dezelfde praktijk (Van Weel en Van Zelst, 1983*) en tussen een ziekenhuispopulatie en een populatie uit de huisartspraktijk (Singh, 1984*; Hayes en Harries, 1984*). In het onderzoek van Singh werd gebruik gemaakt van gematchte populaties, terwijl in dat van Hayes en Harries de patiënten vijf jaar van tevoren aselekt waren verdeeld in een huisartsen- en een ziekenhuisgroep.

In alle tot nu beschreven onderzoeken waarin kwaliteit wordt gemeten op process- en outcomeniveau wordt vrijwel uitsluitend de medisch-technische kant van het handelen van de hulpverleners beoordeeld, en is de status de voornaamste gegevensbron (naast enkele metingen aan de patiënt zelf ter bepaling van de - medische - outcome).

Na deze korte beschrijving van onderzoeken waarin kwaliteit gemeten wordt op twee niveau's zonder dat aandacht wordt besteed aan het verband daartussen, volgt nu de beschrijving van een aantal onderzoeken waarin **process- en outcomematen aan elkaar gerelateerd** worden.

In een onderzoek van McGhan naar het verschil in kwaliteit van voorschrijven bij hypertensie tussen apothekers en huisartsen werd gebruik gemaakt van zowel expliciete als impliciete criteria op process- en outcomeniveau omtrent juist voorschrijfbeleid (Mc Ghan e.a., 1983*).

Bass e.a. trachtten na te gaan of kwaliteit van het process, impliciet beoordeeld door twee artsen op basis van patiëntenstatussen, samenhang met de outcome, zoals deze na één of drie maanden door de patiënt werd gerapporteerd (Bass e.a., 1986*). Het ging hierbij om de volgende aandoeningen: buikklachten, rug- of nekpijn, pijn op de borst, vermoeidheidsklachten, hoofdpijn, oogsymptomen, rectaal bloedverlies. Op deze wijze gemeten, bleek het vroeg verdwijnen van klachten (positieve outcome) meer samen te hangen met patiëntgebonden factoren, onder andere ideeën van de patiënt over de arts-patiënt-relatie, dan met de door de artsen beoordeelde kwaliteit van het process.

In twee andere onderzoeken wordt het verband tussen verschillende meetmethoden op process- en outcomeniveau onderzocht, met tegengestelde resultaten. Door Brook wordt een onderzoek beschreven naar de overeenkomsten tussen de resultaten van vier verschillende meetmethoden van process en één van outcome wat betreft de zorg voor patiënten met urineweginfectie, hypertensie en ulcus duodeni/ventriculi (Brook, 1973). Met alle verschillende methoden werden zeer uiteenlopende resultaten gevonden. Aanbevolen werd in individuele gevallen zowel het process als de outcome te evalueren. Kane e.a. vonden echter een duidelijk verband tussen process, zowel indien impliciete als expliciete criteria werden gehanteerd, en outcome, in termen van functionele status en satisfactie (Kane e.a., 1977*).

Toch wordt over het algemeen aangenomen dat het verband tussen process en outcome, zo het er al is, toch niet erg eenduidig is. Dit is de reden waarom sommige auteurs op process- en outcomeniveau meten

en de gegevens per individuele patiënt samenvoegen. Een voorbeeld hiervan is de zogenaamde **'trajectory'-methode**, ontwikkeld door Williamson en duidelijk beschreven in een onderzoek van Zuckerman e.a. naar de kwaliteit van zorg bij patiënten met gastrointestinale aandoeningen (Williamson, 1971; Zuckerman e.a., 1980*). Deze methode maakt onderscheid in diagnostisch proces, diagnostische outcome, therapeutisch proces en therapeutische outcome en volgt als het ware de patiënt tijdens een ziekte-episode. Voor diagnostisch proces, therapeutisch proces en therapeutische outcome zijn van tevoren expliciete criteria opgesteld, waarbij zowel de medisch-technische als de psychosociale kant aan bod komt. Diagnostische outcome wordt bepaald aan de hand van het oordeel van een onafhankelijke deskundige (die niet op de hoogte is van de door de beoordeelde arts gestelde diagnose). Deze methode, die veelbelovend lijkt, is helaas in de Nederlandse huisartsgeneeskunde minder goed toepasbaar, aangezien de huisarts voornamelijk te maken heeft met klachten, in plaats van diagnoses en het proces van uitgebreide diagnostiek meer ligt op het terrein van de medisch specialist.

Een vergelijkbare methode werd ontwikkeld specifiek voor 'primary care' door Mushlin en Appel (Mushlin en Appel, 1980). Uitgaande van de outcome bij een bepaalde patiënt op een bepaald tijdstip na het eerste spreekuurbezoek, gemeten aan de hand van expliciete criteria, wordt het daaraan voorafgaande proces ('blind') beoordeeld, zowel met expliciete criteria als door een impliciet oordeel. Deze methode is in de eerste plaats ontwikkeld ten behoeve van kwaliteitsbevordering in de eerstelijns, maar lijkt ook zeker voor onafhankelijke kwaliteitsmeting geschikt.

Wordt in de voorgaande benadering het proces geëvalueerd onafhankelijk van of de outcome positief of negatief is, in een tweetal andere onderzoeken wordt **uitgegaan van bepaalde negatieve outcomes** als indicatoren om het voorafgaande proces te evalueren. Kramer ging uit van negatieve outcomes bij een aantal omschreven ziektebeelden bij kinderen die voorkomen hadden kunnen worden (bepaalde complicaties, niet tijdige diagnose, therapie of verwijzing) (Kramer, 1984*). Van een groep van 103 'case'-kinderen met negatieve outcomes en 103 gematchte controle-kinderen met dezelfde ziektebeelden bij wie geen complicaties waren opgetreden en wel tijdig diagnose therapie of verwijzing was gerealiseerd, en die allen minstens 24 uur tevoren een arts hadden bezocht, werd retrospectief het proces beoordeeld. Dit onderzoek was bedoeld om eventuele verschillen tussen kinderartsen en algemene artsen op te sporen.

Door Heineken e.a. werd gebruik gemaakt van een tweetal lijsten van zogenaamde 'negative indexes of health', opgesteld door Rutstein e.a. (Heineken e.a., 1985*). De eerste lijst is zodanig samengesteld dat bij het vóórkomen van één geval van een aandoening van deze lijst nader onderzoek moet worden gedaan naar het voorafgegane process; bij de tweede lijst dient nader onderzoek naar het process van zorg te worden ingesteld indien meer dan een van tevoren vastgesteld aantal gevallen van een van de aandoeningen van deze lijst voorkomen. Deze laatste methode vereist een goed opgezet informatiesysteem waarin aandoeningen en complicaties zijn vastgelegd; het lijkt daarom in de Nederlandse eerstelijns op het eerste gezicht niet haalbaar. Bovendien bleek in dit onderzoek dat het vóórkomen van de meeste aandoeningen te gering was om uitspraken te kunnen doen over de kwaliteit van zorg in de instelling. Als dat al het geval is, zal het in de eerstelijns helemaal een probleem worden om conclusies te trekken, aangezien de bedoelde aandoeningen daar nog veel zeldzamer zullen zijn.

De gegevens uit deze paragraaf samenvattend, zien we dat in een aantal onderzoeken kwaliteit gemeten wordt op verschillende niveau's, zonder dat dit iets extra's toevoegt, in vergelijking met meetmethoden die op één van beide niveau's meten.

De onderzoeken waarin resultaten van metingen op process- en outcomeniveau vergeleken worden, leiden tot tegenstrijdige conclusies. Een veelbelovende methode is om uit te gaan van outcomes en het bij een bepaalde outcome behorende process te beoordelen. Uitgaan van negatieve outcomes als screening om nader onderzoek te doen naar het process lijkt in de eerstelijns niet haalbaar.

4. ASPECTEN VAN MEETMETHODEN

4.1. Bronnen van informatie

In deze paragraaf wordt één bepaald onderdeel van de meetmethoden centraal gesteld, namelijk de gehanteerde bronnen van informatie. De aspecten van de mogelijke informatiebronnen waarnaar vooral gekeken wordt, zijn betrouwbaarheid, validiteit en praktische haalbaarheid. De meest gehanteerde bron van informatie die we zijn tegengekomen in dit literatuuroverzicht is de **patiëntenstatus**, in de Nederlandse huisartsgeneeskunde meestal de 'groene kaart'. Dit ondanks het feit dat er nogal wat bezwaren aan kleven betreffende validiteit en betrouwbaarheid. Allereerst is de status geen weergave van wat zich in het contact tussen arts en patiënt afspeelt. Wordt er al genoteerd dat een bepaalde handeling verricht wordt, dan is er nog niet uit op te maken hoe die verricht wordt. Bovendien wordt een aantal 'niet-medische' aspecten van het arts-patiënt-contact überhaupt vrijwel nooit vastgelegd, zoals bijvoorbeeld attitude, empathie en systematisch handelen. Daarnaast is de verslaglegging zeer eenzijdig: afkomstig van de arts, dus gekleurd door zijn/haar oordeel. Wat betreft de betrouwbaarheid, is algemeen bekend dat zelden de vragen genoteerd worden die gesteld zijn, maar waarop de patiënt ontkennend geantwoord heeft. In de (Amerikaanse) ziekenhuissituatie blijkt een relatie tussen de kwaliteit van de medische registratie en de kwaliteit van het medisch handelen wel aantoonbaar (Lyons en Payne, 1974), maar in de eerstelijns lijkt de validiteit en betrouwbaarheid problematisch te zijn. Door Romm en Putnam werd de informatie in statussen vergeleken met woordelijke weergaven van spreekuurcontacten in 'primary care' op basis van audio-opnamen. Slechts 59% van de informatie-items uit één van beide bronnen werd teruggevonden in de status (Romm en Putnam, 1981*). Door Mansfield werden de statussen van huisartsen die werden opgestuurd bij overdracht naar een andere praktijk beoordeeld. De enige items die in alle statussen waren terug te vinden, waren leeftijd en geslacht; de overige gegevens, zowel demografische als medische, ontbraken in een groot aantal gevallen (Mansfield, 1986*). In dit geval is nog afgezien van de vraag of de aanwezige gegevens juist waren.

Thompson en Osborne gingen na of statussen voldoende informatie bevatten om de kwaliteit van zorg te beoordelen aan de hand van expliciete criteria (Thompson en Osborne, 1976*). Nagegaan werd hoeveel

van de 144 informatie-items die nodig waren voor het onderzoek, terug te vinden waren in het merendeel van de 10.500 statussen. 'Het merendeel' was geoperationaliseerd als 75% of meer. Slechts 24 items bleken voor te komen in minstens 75% van alle statussen. Bovendien was slechts 48% van de 166 artsen van oordeel dat de methode een juist beeld gaf van hun handelen.

Voor de Nederlandse situatie maken De Haen en Wijkkel melding van problemen bij het verkrijgen van informatie van de groene kaarten die zij nodig hadden voor hun toepassing van de tracertechniek (De Haen en Wijkkel, 1982). Behalve dat de betrouwbaarheid van de verslaglegging in de status twijfelachtig is, dient ook rekening te worden gehouden met problemen van betrouwbaarheid van het onttrekken van informatie aan statussen. In veel onderzoeken wordt gewerkt met 'abstractors', maar hoe betrouwbaar hun werk is, wordt zelden getest of beschreven.

In een tweetal onderzoeken worden vier verschillende methoden van informatie verzamelen onderling vergeleken wat betreft betrouwbaarheid, validiteit, kosten en aanvaardbaarheid voor arts en patiënt. Het gaat hierbij om onderzoek naar de kwaliteit van hulpverlening bij poliklinische zorg voor patiënten met chronisch obstructieve longaandoeningen (Kemeny e.a., 1984*; Gerbert en Hargreaves, 1986*). Deze vier methoden zijn: een **interview met de arts** na afloop van het consult, een **interview met de patiënt** na afloop van het consult, de **patiëntenstatus** en **video-opnamen**. De betrouwbaarheid van alle vier de methoden bleek hoog (groter dan 0,80); de validiteit was het beste bij de beide interviewmethoden. Om optimaal resultaat te krijgen, wordt aanbevolen twee of meer methoden tegelijkertijd te gebruiken, waaronder in ieder geval een interviewmethode.

In een aantal onderzoeken wordt gebruik gemaakt van de methode van '**papieren patiënten**'. De te onderzoeken artsen krijgen hierbij een schriftelijke casus voorgelegd, waarop zij kunnen reageren door het stellen van vragen en het doen van onderzoek (op papier). Uiteindelijk wordt een beleid geformuleerd. De normale werkwijze wordt hierbij zoveel mogelijk nagebootst. Het gedrag van de arts wordt beoordeeld op grond van de gemaakte keuzes en genomen beslissingen. Voordeel van deze methode is dat alle onderzochte artsen met dezelfde 'patiënt' worden geconfronteerd. Een nadeel is, dat het handelen van de arts in een kunstmatige situatie wordt beoordeeld, die mogelijk leidt tot anderssoortig gedrag dan in de werkelijkheid. Dit probleem wordt aangeduid met de term 'performance gap'.

Door Kuyvenhoven e.a. werden papieren patiënten gehanteerd om te beoordelen of huisartsen voldoende aandacht besteden aan relevante so-

matische aspecten en oorzaken van de gepresenteerde (vage) klachten en in hoeverre hun beleid kans op onnodige iatrogene schade induceert (Kuyvenhoven e.a., 1984a; Jacobs e.a., 1984, Kuyvenhoven e.a., 1986*). Deze beide aspecten die (impliciet) beoordeeld werden door huisartsen-deskundigen, bleken redelijk betrouwbaar te meten bij drie simulatiecasussen (respectievelijk $r = 0,83$ en $r = 0,89$). Wat betreft induceren van kans op onnodige schade, was de variatie tussen de beoordeelde huisartsen groter dan de variatie per huisarts, hetgeen impliceert dat hier een eigenschap van de huisarts wordt gemeten. Bovendien bleek er een hoge correlatie te bestaan tussen de score op 'gevaar voor onnodige schade' en het bij voortduring in ruime mate voorschrijven van medicijnen uit de categorieën analgetica en antirheumatica, hypnotica/sedativa en tranquillizers, neuroleptica en antidepressiva.

Door Haaijer-Ruskamp e.a. werd de geneigdheid tot en rationaliteit van voorschrijven van 118 huisartsen beoordeeld aan de hand van schriftelijke patiëntensimulaties. Ook hier bleek het voorschrijfgedrag van de huisartsen bij de simulaties samen te hangen met het werkelijke voorschrijfgedrag volgens ziekenfondsgegevens (Haaijer-Ruskamp e.a., 1983*; Haaijer-Ruskamp, 1984).

Dit zijn allemaal indicaties, dat wat gemeten wordt met deze papieren-patiënten-methode samenhangt met het gedrag van huisartsen in de werkelijkheid.

Een poging om de performance gap te elimineren en toch de te beoordelen artsen allemaal met dezelfde patiënt te confronteren, komt tot uiting in het werk van Norman e.a. en McClure e.a. (Norman e.a., 1985*; Woodward e.a., 1985; McClure e.a., 1985). In dit type onderzoek trachten goed getrainde simulatie- of gestandaardiseerde 'echte' patiënten ongemerkt binnen te komen in de praktijken van family physicians, waarna zij het handelen van de artsen vastleggen op zogenaamde 'recallformulieren'. In het onderzoek van Norman e.a. werd een interraterbetrouwbaarheid (K) van 0,86 gevonden. In het onderzoek van McClure e.a. werden de gestandaardiseerde patiënten pas op family physicians (26) 'losgelaten' wanneer ze na training een interraterbetrouwbaarheid van 85 hadden bereikt. In dat onderzoek bleek dat 85% van de gevallen de onderzochte arts verder de registratie 'accuraat' was. Ook bleek 88% van de artsen de patiënt niet als een gestandaardiseerde patiënt te hebben herkend. De vastgelegde gegevens werden vervolgens beoordeeld aan de hand van tevoren vastgestelde expliciete criteria.

Een soortgelijk onderzoek werd in Nederland uitgevoerd door Rethans en Van Boven. Behalve dat het handelen van de huisartsen door de si-

mulatiepatiënten die urineweginfectie simuleerden, werd beoordeeld, werd dezelfde huisartsen circa twee maanden later een vragenlijst toegestuurd over dezelfde casus. Het handelen werd beoordeeld aan de hand van een protocol van het Nijmeegs Universitair Huisartsen Instituut (zie paragraaf 4.2). Opvallend was dat de huisartsen in werkelijkheid meer van de obligate handelingen verrichtten en minder overbodige of facultatieve handelingen dan zij aangaven bij beantwoording van de vragenlijst. Geconcludeerd werd, dat de methode van simulatiepatiënten meer valide was dan de vragenlijstmethode (Rethans en Van Boven, 1987*; 1988).

De bronnen van informatie waarbij het gevaar van beïnvloeding door de meetmethode optreedt, zijn: directe observatie, papieren patiënten, video-opnames, interview- en vragenlijstmethoden. Het probleem van de 'performance gap' is echter niet bij alle genoemde methoden even groot. Bij de meeste onderzoeken waarbij artsen onderling vergeleken worden (en niet 'absolute kwaliteit' gemeten wordt) is dit dan ook geen doorslaggevend argument om bijvoorbeeld geen gebruik te maken van video-opnames. Het probleem van beïnvloeding van de resultaten door de meetmethode zelf bestaat niet bij informatie die al vóór het onderzoek was vastgelegd, zoals informatie in de status of 'produktiecijfers' over voorschrijven en verwijzen, en evenmin bij de methode van simulatiepatiënten, aangezien de hulpverlener in al deze gevallen niet op de hoogte is van het feit dat zijn handelen beoordeeld wordt.

De laatste mogelijke informatiebron die wij nog willen beschouwen, is die van de 'produktiecijfers'. In Nederland worden hiervoor vaak ziekenfondscijfers over voorschrijven en verwijzen gehanteerd. Behalve dat hiermee slechts een beeld wordt verkregen van een deel van de 'produktie' van een huisarts, en dan nog voor een geselecteerde groep, namelijk ziekenfondspatiënten, kan men zich afvragen in hoeverre zuiver kwantitatieve gegevens een valide maat kunnen zijn voor de kwaliteit van het handelen van huisartsen. Pogingen die gedaan worden om deze gegevens althans enigszins kwalitatieve waarde te geven, zijn bijvoorbeeld gedaan in het onderzoek van Haaijer-Ruskamp en in dat van Mekkink. In deze beide onderzoeken werd gekeken naar het voorschrijven van bepaalde groepen geneesmiddelen, waarbij met name veel voorschrijven van symptomatische middelen en zogenaamde 'minor tranquillizers' als een teken van kwalitatief minder goed voorschrijven kan worden gezien (Haaijer-Ruskamp, 1984; Mekkink, 1986).

Nog een probleem bij voorschrijfcijfers is, dat niet wordt vastgelegd of een bepaald voorschrift een eerste, dan wel een herhaalre-

cept betreft. In Engeland werd een nieuw systeem van receptenbriefjes onderzocht, teneinde meer specifieke informatie hieromtrent te verkrijgen ten behoeve van feed-back aan de huisartsen zelf (Fraser en Gosling, 1985). Nog een mogelijkheid zou zijn om per recept tevens vast te leggen wat volgens de voorschrijvende huisarts de indicatie voor dit recept is. Voorzover ons bekend, is deze methode nog nooit uitgeprobeerd.

Voor ruwe verwijscijfers geldt in grote lijnen hetzelfde als voor voorschrijfcijfers; ook hier zou vastlegging van de reden van verwijzing meer aanknopingspunten bieden voor onderzoek naar kwaliteit dan zuiver kwantitatieve gegevens.

Door Given e.a. wordt het gebruik van een uitgebreid informatiesysteem beschreven om 'primary care residents' in een kliniek te beoordelen. Naar het handelen van 'residents' wier produktiecijfers ver boven of ver onder het gemiddelde liggen, wordt nader onderzoek ingesteld. Ook hier wordt het probleem gesignaleerd dat het absoluut niet zeker is, dat een arts in opleiding met gemiddelde kwantitatieve produktiecijfers, ook inhoudelijk goede zorg levert (Given e.a., 1981*).

Geautomatiseerde gegevensverzameling lijkt een veelbelovende ontwikkeling voor de toekomst. Indien het mogelijk wordt gedetailleerde informatie van huisartsen omtrent aandoeningen en handelingen routinematig in te voeren in een informatiesysteem, dan betekent dat een belangrijke verbetering ten opzichte van de huidige routinematig vastgelegde gegevens. Bij de Nationale Studie naar Ziekten en Verichtingen in de Huisartspraktijk van het NIVEL is steeds gedurende drie maanden zeer gedetailleerde informatie door huisartsen verstrekt. Een systeem waarbij dit soort informatie routinematig wordt ingevoerd, heeft als voordelen dat het veel toegankelijker en gedetailleerder is dan 'de groene kaart', terwijl het toch gaat om gegevens die routinematig worden vastgelegd, dus geen probleem van 'performance gap'. Het zal echter wel de nodige problemen geven om een dergelijk systeem blijvend in te voeren en te bewerkstelligen dat de ingevoerde informatie betrouwbaar is.

Een aspect van de informatiebron dat nog nauwelijks ter sprake is gekomen, is de praktische haalbaarheid. Het is duidelijk dat de praktische haalbaarheid van het gebruik van statussen en ziekenfondscijfers veel groter is dan van observaties en simulatiepatiënten. Naast het methodologische voordeel van papieren en 'levende' simulaties, dat alle artsen in dezelfde situatie worden beoordeeld, vormt de investering die nodig is om een casus uit te werken - zowel qua inspanning als qua kosten - een groot praktisch struikelblok.

Tot nu toe hebben wij bij de beschouwing van de validiteit van de informatiebronnen buiten beschouwing gelaten wat men verstaat onder het begrip kwaliteit of van welk onderdeel van het handelen van huisartsen men kwaliteit wil meten. Dit is afhankelijk van de opzet en het doel van het onderzoek: voor het beoordelen van alleen de kwaliteit van de verslaglegging door huisartsen (als onderdeel van het organisatorisch handelen) is de groene kaart een goede informatiebron; ziekenfondscijfers omtrent verwijzen kunnen zeer bruikbaar zijn als eerste 'screening', voorafgaand aan een diepergaand onderzoek naar de kwaliteit van voorschrijfgedrag. Bestudering van de arts-patiënt-relatie is goed mogelijk op basis van audio- of video-opnamen, maar wil men bijvoorbeeld het diagnostisch handelen van huisartsen gedetailleerd bestuderen (dat wil zeggen inclusief lichamelijk onderzoek), dan is directe observatie vrijwel onontbeerlijk. Tenslotte is het behalve voor de keuze van de bron van informatie uiteraard ook voor de wijze van de toepassing van de meetmethode en de onderzoeksofzet belangrijk rekening te houden met het doel van het onderzoek. Over welke groep wil men uitspraken doen, hoe generaliseerbaar moeten de conclusies zijn; is het een inventarisatie of een vergelijkend onderzoek? Dergelijke vragen dienen uiteraard ook in ogenschouw te worden genomen bij de opzet van onderzoek.

Samenvattend kunnen wij over de bronnen van informatie het volgende zeggen (zie bijvoorbeeld ook Barro, 1973): de patiëntenstatus wordt veel gebruikt, is een praktisch goed bruikbare bron van informatie; de informatie was meestal al van te voren vastgelegd, wordt dus niet beïnvloed door de meetmethode, echter validiteit en betrouwbaarheid zijn niet optimaal. De interviewmethode bij arts of patiënt wordt in een tweetal onderzoeken beschreven als betrouwbaar en valide. Het is echter een zeer kostbare methode, en bovendien wordt naar onze mening de validiteit aangetast door de mogelijkheid dat vooral sociaal wenselijke informatie wordt verstrekt. Video- of audio-opnamen geven een goede mogelijkheid op betrouwbare wijze informatie te verzamelen die voor bepaalde aspecten van het handelen van huisartsen ook valide zijn; er is echter wel het gevaar van beïnvloeding van het gedrag van arts en/of patiënt. Dit laatste is ook een nadeel van de directe observatie naast het probleem van de praktische haalbaarheid. Voor simulatiepatiënten geldt niet het bezwaar van beïnvloeding; validiteit en betrouwbaarheid zijn hoog, de haalbaarheid is echter matig, met name als men verschillende casussen zou willen gebruiken, ten einde een meer omvattend beeld van het handelen van huisartsen te verkrijgen. Ook bij papieren patiënten is de haalbaarheid beperkt,

bovendien wordt het gedrag van de hulpverlener niet in de normale situatie beoordeeld; de betrouwbaarheid is echter hoog. Productiecijfers zijn relatief gemakkelijk te verkrijgen, liggen meestal voor het onderzoek al vast en worden dus niet door de meetmethode beïnvloed; de validiteit is echter verre van optimaal. Bij de keuze van een informatiebron hangt het af van het doel van het onderzoek welk gewicht aan bepaalde aspecten wordt toegekend. Het is niet mogelijk één bron van informatie aan te wijzen die in alle situaties de beste is.

4.2. Criteria

In de door ons aangehouden definitie van kwaliteit speelt de term 'criteria' een belangrijke rol ('de mate van overeenstemming tussen de werkelijkheid en tevoren vastgestelde criteria'). Wat zijn criteria nu precies? Donabedian verstaat onder 'criteria': componenten van zorg die te maken hebben met de kwaliteit van zorg. Men kan criteria beschouwen als operationalisatie van wat volgens de onderzoekers 'de wenselijkheid' is. De term 'criteria' wordt vaak gebruikt als een omvattend begrip voor de drie verwante begrippen criteria, standaarden en normen. Een standaard is volgens Donabedian een precieze kwantitatieve specificatie van een criterium die een bepaalde maat van kwaliteit aangeeft, en met normen wordt bedoeld een algemeen idee van wat goed is aan zorg (Donabedian, 1986). Deze termen worden vaak door elkaar gebruikt; om verwarring te voorkomen, hanteren wij hier de samengestelde term 'criteria'.

Criteria kunnen geformuleerd zijn op het niveau van structure, process of outcome; inhoudelijk kunnen ze te maken hebben met medisch-technische, psychosociale of organisatorische aspecten van de hulpverlening. Een heel belangrijk aspect van criteria, dat we al eerder zijn tegengekomen in paragraaf 3.1, is of ze impliciet of expliciet zijn. Door Donabedian zijn de voor- en nadelen van het gebruik van impliciete en expliciete criteria nog eens duidelijk naast elkaar gezet (Donabedian, 1981). Voordelen van het gebruik van expliciete criteria zijn dat ze betrouwbaar en consistent zijn en eenvoudig te hanteren; het hanteren van expliciete criteria is goedkoper dan beoordeling door deskundigen. Het opstellen en hanteren van expliciete criteria heeft bovendien een zekere kwaliteitsbevorderende werking: discussies over wetenschappelijke validiteit en sociale aanvaardbaarheid van het professionele handelen worden aangemoedigd; men gaat zich richten naar de bekende standaarden. Het gevaar bestaat

echter dat door het gebruik van expliciete criteria de validiteit van het begrip kwaliteit wordt aangetast, dat 'kwaliteit' gereduceerd wordt tot een ongenueanceerd technisch begrip. Het streven naar maximale validiteit plus maximale betrouwbaarheid leidt dan tot maximale trivialiteit (Smits e.a., 1985*). De meest valide maat blijft het impliciete oordeel. Donabedian is er daarom voorstander van expliciete criteria te hanteren als screeningsinstrument en vervolgens in een diepergaand onderzoek kwaliteit te meten met behulp van impliciete beoordeling.

Een belangrijke vraag bij het vaststellen van kwaliteitscriteria is: door wie worden criteria vastgesteld? Bij het gebruik van impliciete criteria betreft dit de vraag: wie beoordeelt het handelen van de hulpverlener? Het kan bijvoorbeeld nogal wat uitmaken of de kwaliteit van het handelen beoordeeld wordt door een specialist op een bepaald gebied, door een collega, of door een patiënt. Ook bij het vaststellen van expliciete criteria is er een aanzienlijk verschil te verwachten tussen criteria die op verschillende wijzen zijn vastgesteld. Bij serieuze pogingen kwaliteit te meten, wordt gestreefd naar het gebruik van gevalideerde criteria. Aangezien er een groot tekort is aan onderzoeksresultaten waaruit de validiteit van criteria zou kunnen worden afgeleid (bedoeld is onderzoek naar 'efficacy', werkzaamheid, door Williamson gedefinieerd als 'the extent to which a health care intervention can be shown to be beneficial under optimal conditions of care'), wordt vaak volstaan met het vaststellen van criteria op basis van consensus. Ook dan kan het nogal wat uitmaken wie er aan de consensusbepaling deelnemen. Door Brook e.a. werden in het Rand-experiment criteria vastgesteld op basis van consensus van een panel van inhoudelijk deskundigen, voorafgegaan door literatuuronderzoek. Het gebrek aan wetenschappelijke informatie werd hierbij als probleem vermeld. Daarnaast was men achteraf van mening dat het aanbeveling zou verdienen in het vervolg ook patiënten bij de consensusvorming te betrekken in verband met de geringe aandacht voor psychosociale aspecten, zowel van het panel als in de literatuur (Brook e.a., 1977).

De criteria die zijn vastgelegd in de protocollen van het Nijmeegs Universitair Huisartsen Instituut zijn vastgesteld na discussie met deelnemers aan toetsingsbijeenkomsten naar aanleiding van resultaten van literatuuronderzoek. Hierbij werden de criteria dus bepaald door de beroepsgroep zelf en niet door zeer gespecialiseerde deskundigen. Hoewel deze criteria door de beroepsgroep zelf zijn vastgesteld, blijkt toch - ook na een uitgebreid programma van 'quality assurance' (onderlinge toetsing) - lang niet volledig aan de crite-

ria te worden voldaan (Grol, 1987*). Ook in ander onderzoek worden vraagtekens geplaatst bij de juistheid van de criteria in verband met het lage percentage criteria waaraan wordt voldaan (Rethans en Van Boven, 1988). In Amerikaans onderzoek wordt gebruik gemaakt van zogenaamde 'minimumnormen' (Hulka e.a., 1976*). Hierbij blijkt echter dat alle artsen relatief hoog scoren, en de methode dus onvoldoende gevoelig is om verschillen tussen artsen aan te tonen.

Een oplossing voor dit dilemma van optimale criteria versus minimale criteria zou kunnen zijn om 'gemiddelde' criteria vast te stellen door het handelen van huisartsen in de realiteit te onderzoeken. Een eerste aanzet hiertoe is een onderzoek van Lawler en Hosokawa, die aan een aselechte steekproef van 'primary care' artsen vragenlijsten stuurden met casusbeschrijvingen om na te gaan in hoeverre deze artsen overeenkomstige normen hanteerden (Lawler en Hosokawa, 1987). Door de lage response rate (34%) kunnen aan dit onderzoek echter geen vergaande conclusies worden verbonden. Behalve de vraag hoe zinvol bepaalde criteria zijn om bruikbare onderzoeksresultaten te verkrijgen, is het ook belangrijk aandacht te besteden aan de aanvaardbaarheid van verschillende toepassingen van criteria. Behalve dat criteria die vastgelegd zijn door deskundigen weinig gevoelig zijn in het onderscheiden van kwaliteitsverschillen, kunnen dergelijke criteria ook zeer demotiverend werken op de onderzochte hulpverleners. Een belangrijk aspect van kwaliteitsmeting, dat nog niet aan de orde is geweest in dit overzicht, is immers de aanvaardbaarheid voor de hulpverlener wiens handelen onderzocht wordt. Zonder de medewerking van deze meest betrokken groep kan in het geheel geen onderzoek naar kwaliteit worden gedaan. Indien de indruk gewekt wordt dat het eigenlijke doel van de kwaliteitsmeting een poging is om controle uit te oefenen op het handelen van hulpverleners van hogerhand, in plaats van het bevorderen van het professionele handelen, zal hierdoor de bereidheid om medewerking te verlenen ernstig geschaad worden, hetgeen zeker niet de bedoeling van de onderzoekers kan zijn.

Voor het ontwikkelen en vastleggen van criteria voor de kwaliteit van huisartsgeneeskundig handelen, stelt het Nederlands Huisartsen Genootschap een structuur voor in de nota 'Naar criteria voor kwaliteit' (NHG, 1987). Het ontwikkelen en documenteren van kwaliteitscriteria zou dienen te geschieden onder auspiciën van drie commissies. De bedoeling hiervan is te komen tot een uitgebreid bestand van zoveel mogelijk gevalideerde criteria ten behoeve van onderzoek naar en bevordering van de kwaliteit van het handelen van Nederlandse huisartsen.

5. BESCHOUWING EN AANBEVELINGEN

Wanneer we alle verzamelde onderzoeken overzien, valt op dat het merendeel van de onderzoekers heeft getracht kwaliteit te meten op het niveau van process. Over de voors en tegens van kwaliteit meten op processniveau is veel geschreven (onder andere Donabedian, 1966, 1980; Mc Auliffe, 1979). Volgens Donabedian vormt de outcome de uiteindelijke validiteitsmaat voor de kwaliteit van de hulpverlening. Dit wil echter niet zeggen dat het niet uitmaakt of het process van goede of van slechte kwaliteit is. Bovendien is, zeker in de huisartsgeneeskunde, het meten van outcome geen eenvoudige zaak en soms ook niet relevant. De meeste klachten waarmee de huisarts geconfronteerd wordt, zijn vaag en self-limiting van aard, en zelfs bij 'harde' diagnoses is gebleken dat outcomes niet een rechtlijnig gevolg zijn van process (zie paragraaf 3.2 en 3.3). Een mogelijkheid om de validiteit van de meting te vergroten ten opzichte van meting alleen op processniveau zou kunnen zijn een screening op outcomeniveau, in de Nederlandse situatie bijvoorbeeld op grond van productiecijfers, gevolgd door een meer intensieve beoordeling van het voorafgegane process, vergelijkbaar met de onderzoeken van Mushlin en Appel, Kramer en Heineken e.a. (Mushlin en Appel, 1980; Kramer, 1984*; Heineken e.a., 1985*). Deze methode werd ook voorgesteld door de Werkgroep Toetsing van het Nederlands Huisartsen Genootschap (1980). Verder zal hopelijk het initiatief van het NHG met betrekking tot het standaardenbeleid leiden tot nader onderzoek ter validering van criteria.

Bij de beschouwing van de literatuur is gebleken dat er nogal wat variatie is tussen verschillende onderzoeken wat betreft de aspecten van het handelen van huisartsen die beoordeeld worden, zowel wat betreft aard als omvang. In sommige onderzoeken wordt alleen de kwaliteit van het voorschrijven bij een bepaalde aandoening geëvalueerd, terwijl in andere onderzoeken zowel een groot aantal medisch-technische handelingen als psychosociale aspecten beoordeeld worden. Beschouwt men het begrip 'kwaliteit van het handelen van de hulpverlener' als zijnde opgebouwd uit een groot aantal deelaspecten, waarvan 'kwaliteit van het medisch-technisch handelen' en 'kwaliteit van het psychosociaal handelen' twee belangrijke categorieën zijn naast mogelijke andere categorieën, zoals 'kwaliteit van de organisatie' - vergelijk de discussienota van de Nationale Raad voor de Volksgezondheid (1986) - dan geeft de meer omvangrijke benadering

een meer valide beeld van het begrip 'kwaliteit' dan een benadering van één enkel deelaspect.

In slechts weinig onderzoeken wordt de kwaliteit van het medisch-technisch handelen en die van het psychosociaal handelen beide als gelijkwaardige aspecten beschouwd. Er lijkt een kloof te bestaan tussen de onderzoeken naar deze verschillende aspecten, waarbij het opvallend is dat in onderzoek naar medisch-technische aspecten de patiënt een zeer ondergeschikte rol krijgt toebedeeld. In dit soort onderzoek wordt de patiënt meestal beschouwd als een soort lijdend voorwerp dat het hulpverleningsproces ondergaat en waaraan outcomes gemeten kunnen worden. Dit wordt ook enigszins gereflecteerd in het model van structure, process en outcome, waarin naar onze mening de 'input' van de patiënt ontbreekt. Naar onze mening is het niet terecht dat de invloed van de patiënt op zowel process als outcome in onderzoek zo weinig aandacht krijgt. Uit het in paragraaf 3.3 beschreven onderzoek van Bass e.a. bleek bijvoorbeeld de outcome meer samen te hangen met patiëntgebonden factoren dan met de kwaliteit van het process, zoals dat door artsen werd beoordeeld. Ook wat betreft het vaststellen van criteria, met name voor niet-medische aspecten van kwaliteit, lijkt het ons aanbevelenswaardig naar meer inbreng van patiënten te streven. Met de beroepsgroep zijn zij immers het meest betrokken bij de kwaliteit van de zorg. Men dient echter ook rekening te houden met een aantal tekortkomingen bij het gebruik van het oordeel van patiënten. Tevredenheid van de patiënt op zich is bijvoorbeeld zeker geen valide maat voor kwaliteit van de hulpverlening. Met de inbreng van de patiënt dient daarom niet onoordeelkundig te worden omgesprongen. Het door de geneeskundige hoofdinspectie voorgestelde certificatie-instituut, waarin een grote verscheidenheid aan belangengroeperingen zitting zou moeten hebben (Van den Boogaard, 1987), draagt naar ons idee te veel het gevaar van afschrikking van de beroepsbeoefenaren in zich, uit angst voor bovenmatige bemoeienis van overheid en financiers. Ons lijken beoordelingsinstrumenten, zoals die bijvoorbeeld door het NUHI ontwikkeld zijn, die zowel het medisch-technisch handelen als psychosociale aspecten omvatten, een goed uitgangspunt om voor de Nederlandse situatie valide criteria voor de kwaliteit van het handelen van huisartsen te ontwikkelen. Landelijke supervisie en documentatie, zoals door het NHG voorgesteld, kan de coördinatie en continuïteit van onderzoek zeer ten goede komen. Aanbevolen wordt, dat de huidige meetinstrumenten getest worden in een aselekt samengestelde groep huisartsen, en dat de criteria op basis van dit onderzoek bijgesteld worden tot meer 'realistische' criteria. Daarnaast verdient het aan-

beveling onderzoek te doen waarin deze criteria op processniveau getoetst worden aan outcome-maten (validering) en ook, dat bij het verder ontwikkelen van criteria meer aandacht wordt geschonken aan de mening van patiënten.

Het ontwikkelen van meetmethoden van kwaliteit vereist dat de verschillende onderzoekers bij de opzet van hun onderzoek rekening houden met eerder verricht onderzoek en daar zoveel mogelijk op aansluiten. Gezien de omvangrijkheid van de literatuur op dit gebied, is het niet eenvoudig een dergelijke continuïteit van onderzoek te handhaven.

Hopelijk kan dit literatuuroverzicht een bijdrage leveren aan het bevorderen van die noodzakelijke continuïteit.

GERAADPLEEGDE LITERATUUR

- ANDERSEN, O.W. en M.C. SHIELDS. Quality measurement and control in physician decision making: state of the art. *Health Services Research*; 17, 1982, nr. 2, p. 125-155
- MC AULIFFE, W.E. Measuring the quality of medical care: process versus outcome. *Milbank Memorial Fund Quarterly*; 57, 1979, nr. 1, p. 118-152
- BARRO, A.R. Survey and evaluation of approaches to physician performance measurement. *Journal of Medical Education*; 48, 1973, nr. 12, p. 1051-1093
- BASS, M.J., C. BUCK, L. TURNER, G. DICKIE, G. PRATT, H.C. ROBINSON. The physician's actions and the outcome of illness in family practice. *The Journal of Family Practice*; 23, 1986, nr. 1, p. 43-47
- BECHT-MELAI, F. Registratie wetenschappelijk onderzoek eerstelijnsgezondheidszorg 1986, nr. 15. Utrecht: NIVEL, 1987
- BENNETT, S.E., J.O. GOODSON, J.E. IZEN, W.T. BRANCH, W.C. CLARK, C.J. HATEM, R.S. LAWRENCE, T.L. DELBANCO, A.H. GOROLL. Comparing ambulatory care practices of primary care and traditional medicine residents. *Medical Care*; 23, 1985, nr. 6, p. 816-822
- BERGNER, M. Measurement of health status. *Medical care*; 23, 1985, nr. 5, p. 696-704
- BERGNER, M., R.A. BOBBITT, W.B. CARTER, B.S. GILSON. The sickness impact profile: development and final revision of a health status measure. *Medical Care*; 19, 1981, nr. 8, p. 787-805
- BERGNER, M., R.A. BOBBITT, S. KRESSEL, W.E. POLLARD, B.S. GILSON, J.R. MORRIS. The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Services*; 6, 1976, nr. 3
- BOGAARD, C.M.J. VAN DEN. Kwaliteit huisartsgeneeskunde: mogelijkheden om te komen tot een garantie voor kwaliteit van de door de huisarts verleende zorg. Rijswijk: Geneeskundige Hoofdinspectie van de Volksgezondheid, 1988
- BOLLAM, M.J., M. MC CARTHY, M. MODELL. Patients' assessment of out of hours care in general practice. *British Medical Journal*; 296, 1988, p. 829-832
- BORGIEL, A., J. WILLIAMS, G. ANDERSON et al. Assessing the quality of care in family physicians practices. *Canadian Family Physician*; 31, 1985, p. 853-861

- BRIDGSTOCK, M. Social theory and measures of the quality of medical care in general practice. *Social Science and Medicine*; 1979, p. 269
- BROOK, R.H. Quality of care assessment: a comparison of five methods of peer review. Rockville, Md. 1974. Department of Health, Education and Welfare, Publication no. (HRS) 74-3100
- BROOK, R.H., K.N. LOHR, Efficacy, effectiveness, variations and quality: boundary-crossing research. *Medical Care*; 23, 1985, nr. 5, p. 710-722
- BROOK, R.H., A. DAVIES-AVERY, S. GREENFIELD, L.J. HARRIS, T. LELAH, N.E. SOLOMON, J.E. WARE Jr. Assessing the quality of medical care using outcome measures: an overview of the method. *Medical Care Suppl*; 15, 1977, nr. 9, 165 p.
- BROUWER, W., P.D. KERKHOF, P. POP. Diagnostisch Centrum Maastricht, de relatie 1^e - 2^e lijn in ontwikkeling: evaluatie van het experiment diagnostisch centrum Maastricht. Eindrapportage. Maastricht: Rijksuniversiteit Limburg, 1983, 186 p.
- McCLURE, G.L., E.P. GALL, K.E. MEREDITH, M. ANNELL-GOODEN, J.T. BOYER. Assessing clinical judgement with standardized patients. *Journal of Family Practice*; 20, 1985, nr. 5, p. 457-464
- CREBOLDER, H., M. VAN EYNDHOVEN. De begeleiding van type II diabeten. *Huisarts en Wetenschap*; 29, 1986, p. 16-19
- CRIJNEN, A.A.M., J. VAN THIEL, H.F. KRAAN. Evaluatie van consultvoering: een spreekuur nagebootst. *Huisarts en Wetenschap*; 29, 1986, p. 316-318
- DIMATTEO, M.R., D.D. DINICOLA. Sources of assessment of physician performance: a study of comparative reliability and patterns of intercorrelation. *Medical Care*; 19, 1981, nr. 8, p. 829-842
- DONABEDIAN, A. Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly*; 1966, p. 166-206
- DONABEDIAN, A. The definition of quality and approaches to its assessment: explorations in quality assessment and monitoring. *Health Administration Press Ann Arbor Michigan*; 1, 1980
- DONABEDIAN, A. Advantages and limitations of explicit criteria for assessing the quality of health care. *Milbank Memorial Fund Quarterly*; 59, 1981, nr. 1, p. 99-106
- DONABEDIAN, A. The methods and findings of quality assessment and monitoring: an illustrated analysis. *Explorations in quality assessment and monitoring Vol. III. Health Administrated Press, Ann Arbor Michigan, 1985*
- DONABEDIAN, A. Criteria and standards for quality assessment and monitoring. *Quality Review Bulletin*, 1986, nr. 3, p. 99-108

- DRYER WRIGHT, D., R.L. KANE. Predicting the outcome of primary care. *Medical Care*; 20, 1982, nr. 2
- DUTTON, D.B., D. GOMBY, J. FOWLES. Satisfaction with children's medical care in six different ambulatory settings. *Medical Care*; 23, 1985, nr. 7, p. 894-912
- EDWARDS, M.W., W.M. FORMAN, J. WALTON. Audit of abdominal pain in general practice. *Journal of the Royal College of General Practitioners*; 35, 1985, p. 235-238
- FASSNACHT, G. Theory and practice of observing behaviour. London: Academic Press, 1982
- FLEMING, D.M., M.S.T.A. LAWRENCE. Impact of audit on preventive measures. *British Medical Journal*; 287, 1983, nr. 6408, p. 1852-1854
- FLETCHER, R.H., M.S. O. MALLEY, S.W. FLETCHER, e.a. Measuring the continuity and coordination of medical care in a system involving multiple providers. *Medical Care*; 22, 1984, nr. 5, p. 403-411
- FRASER, R., J.T.L. GOSLING. Information systems for general practitioners for quality assessment: III suggested new prescribing profile. *British Medical Journal*; 29, 1985, nr. 6509, p. 1613-1615
- FREELING, P. Health outcomes in primary care: an approach to the problems. *Family Practice*; 1985, nr. 2, p. 177-181
- GERBERT, B., W.A. HARGREAVES. Measuring physician behavior. *Medical Care*; 24, 1986, nr. 9, p. 838-847
- GERRITSMAN, J.G.M., J.A. SMAL. De werkwijze van huisarts en inter-
nist: een vergelijkend onderzoek met behulp van een interactieve patiëntensimulatie. Utrecht: Wetenschappelijke Uitgeverij Bunge, 1982. Dissertatie Rijksuniversiteit Utrecht, 303 p.
- GEYNDT, W. DE. Five approaches for assessing the quality of care. *Hospital Administration*, 1970, p. 21-42
- MC GHAN, W.F., G.L. STIMMEL, T.G. HALL, T.M. GILMAN. A comparison of pharmacists and physicians on the quality of prescribing for ambulatory hypertensive patients. *Medical Care*; 21, 1983, nr. 4, p. 435-444
- GIVEN, C.W., M. BROWNE, R.J. SPRAFKA, E.C. BRECK. Evaluating primary ambulatory care with a health information system. *The Journal of Family Practice*; 12, 1981, nr. 2, p. 293-302
- GONELLA, J.S., M.J. GORAN. Quality of patient care: a measurement of change: the staging concept. *Medical Care*; 13, 1975, nr. 6, p. 467-473
- GRACE, J.F., D. ARMSTRONG. Referral to hospital: perceptions of pa-

- tients, general practitioners and consultants about necessity and suitability of referral. *Family Practice*; 4, 1987, nr. 3, p. 170-175
- GROL, R. Kwaliteitsbewaking in de huisartsgeneeskunde: effecten van onderlinge toetsing. Dissertatie. Nijmegen: Katholieke Universiteit Nijmegen, 1987
- GROL, R. Preventie van somatische fixatie en de attitude van de huisarts. Dl. 5 eindrapportage. Nijmeegs Universitair Huisartsen Instituut, 1983, 144 p.
- GROL, R., J. VAN EIJK, H. MOKKINK, M. BEEK, P. MESKER, A. SMITS, J. MESKER-NIESTEN. Taakopvatting van de huisarts en zijn handelen in de spreekkamer. *Gezondheid en Samenleving*; 6, 1985, nr. 1, p. 31-40
- GROL, R., H. MOKKINK, A. SMITS, J. VAN EIJK, M. BEEK, P. MESKER, J. MESKER-NIESTEN. Work satisfaction of general practitioners and the quality of patient care. *Family Practice*; 2, 1985, nr. 3, p. 128-135
- GROL, R., H. MOKKINK, F. SCHELLEVIS. Effecten van onderlinge toetsing: I spreekurgedrag van huisartsen. *Huisarts en Wetenschap*; 29, 1986a, nr. 4, p. 99-103
- GROL, R., H. MOKKINK, J. VAN DEURSEN, A. TEUNISSEN. Effecten van onderlinge toetsing: II voorschrijf- en verwijsgedrag van huisartsen. *Huisarts en Wetenschap*; 29, 1986b, nr. 5, p. 141-145
- GROL, R., P. MESKER (red.). Huisarts en onderlinge toetsing: methoden, normen, protocollen, 2^e herziene druk. Utrecht: Bunge, 1986
- GROL, R., H. MOKKINK, F. SCHELLEVIS. The effects of peer review in general practice. *Journal of the Royal College of General Practitioners*; 38, 1988, p. 10-13
- HAAIJER-RUSKAMP, F.M. Het voorschrijfgedrag van de huisarts: een studie bij 118 huisartsen naar factoren die verschillen in voorschrijfgedrag verklaren. Dissertatie. Groningen: Rijksuniversiteit, 1984, 151 p.
- HAAIJER-RUSKAMP, F.M., R. VAN DER LENDE, J. ORMEL, H. WESSELING. Het voorschrijfgedrag van de huisarts: morbo-cartie of ratiocra-tie?: een studie naar geneigdheid tot en rationaliteit van voorschrijven. *Gezondheid en Samenleving*; 4, 1983, nr. 4, p. 246-255
- HAAN, M. DE, H. HOLLENBEEK BROUWER. Normen voor het dokteren van de huisarts: een weinig omschreven vak. *Medisch Contact*; 37, 1982, nr. 51/52, p. 1650-1653
- HAEN, F. DE, D. WIJKEL. De tracermethode in gezondheidszorgonderzoek. *Gezondheid en Samenleving*; 4, 1982, nr. 3, p. 276-282
- HALL, J., N. HALL, E. FISHER, D. KILLER. Measurement of outcomes of

- general practice: comparison of three health status measures. *Family Practice*; 4, 1987, nr. 2, p. 117-122
- HALL, J.A., M.C. DORNAN. Metaanalysis of satisfaction with medical care: a description of research domain and analysis of overall satisfaction levels. *Social Science and Medicine*; 27, 1988, nr. 6, p. 637-644
- HALL, J.A., M.C. DORNAN. What patients like about their medical care and how often they are asked: a metaanalysis of the satisfaction literature. *Social Science and Medicine*; 27, 1988, nr. 9, p. 935-939
- HAWTON, K., S. MC KEOWN, A. DAY, P. MARTIN, M. O'CONNOR, J. YULE. Evaluation of out-patient counselling compared with general practitioner care following overdoses. *Psychological Medicine*; 1987, 17, p. 751-761
- HAYES, T.M., J. HARRIES. Randomised controlled trial of routine hospital clinic care versus routine general practice care for type II diabetes. *British Medical Journal*; 289, 1984, p. 728-730
- HEINEKEN, P.A., G. CHARLES, D.H. STIMSON, C. WENELL, R.H. STIMSON. The use of negative indexes of health to evaluate quality of care in a primary care group practice. *Medical Care*; 23, 1985, nr. 3, p. 198-208
- HILL, J.D., J.R. HAMPTON, J.R.A. MITCHELL. A randomised trial of home versus hospital management for patients with suspected myocardial infarction. *The Lancet*; 1978, p. 837-841
- HOLLAND, W.W. (ed.). *Evaluation of health care*. Oxford University Press; 1984, 268 p.
- HOLMES, CHR., R.L. KANE, M. FORD, J. FOWLER. Toward the measurement of primary care. *Milbank Memorial Fund Quarterly*; 56, 1978, nr. 2, p. 231-252
- HULKA, B.S., L.L. KUPPER, J.C. CASSEL. Physician management in primary care. *American Journal of Public Health*; 66, 1976, nr. 12, p. 1173-1179
- HULKA, B.S., S.J. ZYZANSKI, J.C. CASSEL, S.J. THOMPSON. Scale for the measurement of attitudes towards physicians and primary medical care. *Medical Care*; 8, 1970, nr. 5, p. 429-435
- HULSHOF, R.E., H.J.P.M. DIJKHUIS, J. THEUNISSEN, J.D. MULDER. Beoordeling van consulten in de huisartspraktijk. *Huisarts en Wetenschap*; 25, 1982, nr. 11, p. 431-434
- HUNT, S.M., J. MC EWEN, S.P. MC KENNA. Measuring health status: a new tool for clinicians and epidemiologists. *Journal of the Royal College of General Practitioners*; 35, 1985, p. 185-188
- IMANTS, J. Huisarts en hypertensie, een observatie onderzoek.

- Utrecht: Nederlands Huisartsen Instituut, 1979
- JACOBS, H.M., M.M. KUYVENHOVEN, F.W. TOUW-OTTEN, J.C. VAN ES. Written simulation of patient-doctor encounters: 2 assessment of the performance of general practitioners. *Family Practice*; 1, 1984, nr. 1, p. 20-24
- KANE, R.L., J. GARDNER, D.D. WRIGHT, G. SNELL, D. SUNDWALL, F.R. WOOLLEY. Relationship between process and outcome in ambulatory care. *Medical Care*; 1977, nr. 15, p. 961
- KAPTEIN, A.A., F.W. DEKKER, K. GILL, M.A.C. VAN DER WAART. Undertreatment of asthma in Dutch general practice. *Family Practice*; 4, 1987, nr. 3, p. 219-225
- KAPTEIN, A.A., F.W. DEKKER, M.A.C. VAN DER WAART, K. GILL. Huisarts en astma: werk aan de winkel; resultaten van de eerste studie naar de aard en kwaliteit van de zorg van de huisarts voor astmapatiënten. *Maatschappelijke Gezondheidszorg*; 14, 1986, nr. 5, p. 20-22
- KEKKI, P. Quality of care in general practice. *Scandinavian Journal of Primary Health Care*; 1987, nr. 5, p. 195-200
- KEKKI, P. Quality assurance in primary medical care. *Scandinavian Journal of Primary Health Care*; 1, 1983, nr. 3/4, p. 132-138
- KEMENY, M.E., W.A. HARGREAVES, B. GERBERT, G.C. STONE, D.S. GULLION. Measuring adequacy of physician performance: a preliminary comparison of four methods in ambulatory care of chronic obstructive pulmonary disease. *Medical Care*; 22, 1984, nr. 7, p. 620-643
- KESSNER, D.M. Quality assessment and assurance: early signs of cognitive dissonance. *New English Journal of Medicine*; 298, 1978, nr. 7, p. 381-386
- KESSNER, D.M., C.E. KALK, J. SINGER. Assessing health quality: the case for tracers. *The New England Journal of Medicine*; 288, 1973, nr. 4, p. 189-194
- KIND, P., R. CARR-HILL. The Nottingham health profile: a useful tool for epidemiologists? *Social Science and Medicine*; 25, 1987, nr. 8, p. 905-910
- KNOTTNERUS, J.A., C.W. DE VET. Effectiviteitsonderzoek in de huisartspraktijk. *Huisarts en Wetenschap*; 30, 1987, nr. 4, p. 116-119
- KNOTTNERUS, J.A., H.C.W. DE VET. Effectiviteitsonderzoek in de huisartspraktijk. Algemene inleiding. *Huisarts en Wetenschap*; 30, 1987, p. 49
1. Niet-gerandomiseerd onderzoek. *Huisarts en Wetenschap*; 30, 1987, p. 50-53

2. Gecontroleerd interventie-onderzoek op basis van radomatisatie. *Huisarts en Wetenschap*; 30, 1987, p. 72-75
 3. Algemene opmerkingen over gerandomiseerd interventie-onderzoek in de huisartspraktijk. *Huisarts en Wetenschap*; 30, 1987, p. 116-119
- KRAAN, H.F., A.A.M. CRIJNEN. The Maastricht history-taking and advice checklist. Proefschrift. Amsterdam: Uitgeverij Lundbeck, 1987
- KRAMER, M.S. The use of preventable adverse outcomes to study the quality of child health care. *Medical Care*; 22, 1984, nr. 3, p. 223-230
- KROL, R.A., D.J. NORDLUND. Patient satisfaction data and residents' physician patient skills. *The Journal of Family Practice*; 17, 1983, nr. 1, p. 141-142
- KUIJLEN, H. Bibliografie methoden en technieken binnen gezondheidszorgonderzoek. Utrecht: NIVEL, 1987
- KURJI, K.H., A.P. HAINES. Detection and management of hypertension in general practices in north west London. *British Medical Journal*; 288, 1984, p. 903-906
- KUYVENHOVEN, M.M., H.M. JACOBS, F.W.M.M. TOUW-OTTEN, J.C. VAN ES. Written simulation of patient-doctor encounters: research instrument for registration of the performance of general practitioners (part 1). *Family Practice*; 1, 1984a, nr. 1, p. 14-19
- KUYVENHOVEN, M.M., H.M. JACOBS, F.W.M.M. TOUW-OTTEN, J.C. VAN ES. Comparison of the performance in the simulation with prescription and the referral data in reality. *Family Practice*; 1, 1984b, nr. 2, p. 25-29
- KUYVENHOVEN, M.M., H.M. JACOBS, F.W.M.M. TOUW-OTTEN, J.C. VAN ES. Verschillen in hulpverlening bij patiënten met vage klachten. *Huisarts en Wetenschap*; 29, 1986, nr. 3, p. 65-71
- KUYVENHOVEN, M.M., H.M. PIETERS, H.M. JACOBS, F.W.M.M. TOUW-OTTEN. De kwaliteit van de hulpverlening bij patiënten met vage klachten. In: *De eerstelijns onderzocht*, W.G.W. Boerma en L. Hingstman (red.). Deventer: Van Loghum Slaterus, 1985, p. 40-49
- LAWLER, F.H., M.C. HOSOKAWA. Evaluation of standards of practice for primary care physicians using 12 hypothetical cases. *Journal of Family Practice*; 24, 1987, nr. 4, p. 377-383
- LEEuw, A.K. DE, A. VOS. Bibliografie: communicatie (huis)arts-patiënt; 1, 1986, nr. 8. Utrecht: NIVEL, 1987
- LIKE, R., S.J. ZYZANSKI. Patient satisfaction with the clinical encounter: social psychological determinants. *Social Science and Medicine*; 24, 1987, nr. 4, p. 351-357

- LOGERFO, J., E. LARSON, W. RICHARDSON. Assessing the quality of care for urinary tract infection in office practice: a comparative organizational study. *Medical Care*; 16, 1978, nr. 6, p. 488-495
- LYONS, T.F., B.C. PAYNE. The relationship of physicians' medical recording: performance to their medical care performance. *Medical Care*; 12, 1974, nr. 5, p. 463-469
- MANSFIELD, B.G. How bad are medical records: a review of the notes received by a practice. *Journal of the Royal College of General Practitioners*; 36, 1986, nr. 290, p. 405-406
- MARQUIS, M.S., A.R. DAVIES, J.E. WARE. Patient satisfaction and change in medical care providers: a longitudinal study. *Medical Care*; 21, 1983, nr. 8, p. 821-829
- MAXWELL, J.A., L.J. SANDLOW, P.G. BASHOOK. Effect of a medical care evaluation program on physician knowledge and performance. *Journal of Medical Education*; 59, 1984, p. 33-38
- MEYBOOM-DE JONG, B., T.E. POSTMA, J. VAN DER ENDE, H. LAMBERTS. De functionele toestand van patienten: theoretische overwegingen bij de ontwikkeling van een meetinstrument. *Huisarts en Wetenschap*; 29, 1986a, nr. 1, p. 8-10
- MEYBOOM-DE JONG, B., T.E. POSTMA, J. VAN DER ENDE, H. LAMBERTS. De functionele toestand van patiënten: een proefonderzoek om de functionele toestand vast te stellen. *Huisarts en Wetenschap*; 29, 1986b, nr. 1, p. 11-14
- MICHAEL, G. Quality of care in managing hypertension by case finding in north west London. *British Medical Journal*; 288, 1984, p. 906-908
- MOKKINK, H.G.A. Ziekenfondscijfers als parameter voor het handelen van huisartsen. Meppel: Krips Repro, 1986, 178 p. Proefschrift Katholieke Universiteit Nijmegen
- MOKKINK, H.G.A., V.C.L. TIELENS, A.J.A. SMITS, R.P.T.M. GROL. Werkstijlen van huisartsen: een explorerend onderzoek naar verschillende stijlen van huisartsgeneeskundig handelen. *Huisarts en Wetenschap*; 29, 1986, nr. 3, p. 72-76.
- MORRELL, D.C., M.O. ROLAND. How can good general practitioner care be achieved? *British Medical Journal*; 294, 1987, nr. 6565, p. 161-162
- MURRAY, J.P. A comparison of patient satisfaction among prepaid and fee-for-service patients (published erratum appears in *Journal of Family Practice*; 24, 1987, nr. 5, p. 464). *Journal of Family Practice*; 24, 1987, nr. 2, p. 203-207
- MUSHLIN, A.J., F.A. APPEL. Testing an outcome: based quality assurance strategy in primary care. *Medical Care*; 18, 1980, supple-

ment bij nr. 5

- NATIONALE RAAD VOOR DE VOLKSGEZONDHEID. Discussienota begrippenkader kwaliteit beroepsuitoefening, 1986
- NEDERLANDS HUISARTSEN GENOOTSCHAP. Naar criteria voor kwaliteit: standaardenbeleid NHG. Utrecht: NHG, 1987
- NORMAN, G.R., V.R. NEUFELD, A. WALSH, C.A. WOODWARD, G.A. McCONVEY. Measuring physicians' performance by using simulated patients. *Journal of Medical Education*; 60, 1985, p. 925-934
- NOVICK, L.F., K. DICKINSON, P. ASNES, e.a. Assessment of Ambulatory Care: Application of the Tracer Methodology. *Medical Care*; 14, 1976, nr. 1, p. 1-12
- PALMER, R.H., R. STRAIN, J.V.W. MAURER, J.K. ROTHROCK, M.S. THOMPSON. Quality assurance in eight adult medicine group practices. *Medical Care*; 22, 1984, nr. 7, p. 632-643
- PALMER, R.H., T.A. LOUIS, L.N. HSU, H.F. PETERSON, J.K. ROTHROCK, R. STRAIN, M.S. THOMPSON, E.A. WRIGHT. A randomized controlled trial of quality assurance in sixteen ambulatory care practices. *Medical Care*; 23, 1985, nr. 6, p. 751-770
- PATRICK, D.L., E. SCRIVENS, J.R.H. CHARLTON. Disability and patient satisfaction with medical care. *Medical Care*; 21, 1983, nr. 11, p. 1062-1075
- PAYNE, B.C., T.F. LYONS, E. NEUHAUS. Relationships of physician characteristics to performance quality and improvement. *Health Services Research*; 19, 1984, nr. 3, p. 307-332
- PENDLETON, D., J. HASLER (eds.). *Doctor-patient communication*. London: Academic Press, 1983
- PENDLETON, D., T. SCHOFIELD, M. MARINKER (eds.). *In pursuit of quality: approaches to performance review in general practice*. The Royal College of General Practitioners (London), 1986
- PLAISIER, O. Kijken bij de bureu: observatie tussen (huis-)artsen. *Medisch Contact*; 42, 1987, nr. 17, p. 533-534
- REE, J.W. VAN, W. VAN GERWEN, H. VAN DEN HOOGEN. Interventie bij een verhoogd risico op hart- en vaatziekten 2. Resultaten op langere termijn bij de behandeling van hypertensie. *Huisarts en Wetenschap*; 28, 1985, p. 53-58
- RETHANS, J.J.E., C.P.A. VAN BOVEN. De dokter onderzocht: simulatiepatiënten met mictieklachten op het spreekuur van de huisarts. *Huisarts en Wetenschap*; 31, 1988, p. 3-6
- RETHANS, J.J.E., C.P.A. VAN BOVEN. Simulated patients in general practice: a different look at the consultation. *British Medical Journal*; 294, 1987, nr. 6575, p. 809-812
- ROBERTS, J.G., P. TUGWELL. Comparison of questionnaires determining

- patient satisfaction with medical care. Health Services Research; 22, 1987, nr. 5, p. 637-654
- ROMM, F.J., S.M. PUTNAM. The validity of the medical record. Medical Care; 19, 1981, nr. 3, p. 310-315
- ROYAL COLLEGE OF GENERAL PRACTITIONERS. What sort of doctor?: assessing quality of care in general practice. Report from General Practice 23, 1985
- RUDNICK, K.V., W.P. FRASER, J.C. SIBLEY, V.R. NEUFELD, D.L. SACKETT, B.A. GERRARD. The impact of continuing education on the quality of patient care. Canadian Family Physician; 29, 1983, p. 810-818
- SCHELLEVIS, F., R. GROL. Normen en onderlinge toetsing. Huisarts en Wetenschap; 30, 1987, p. 212-215
- SERVANT, J.B., J.A.B. MATHESON. Videorecording in general practice: the patients do mind. Journal of the Royal College of General Practitioners; 36, 1986, p. 555-556
- SHEPS, S., A. ROBERTSON. Evaluation of primary care in a community clinic by means of explicit process criteria. Canadian Medical Association Journal; 131, 1984, p. 881-906
- SIBLEY, J.C., W.O. SPITZER, K.V. RUDNICK, J.D. BELL, R.D. BETHUNE, D.L. SACKETT, K. WRIGHT. Quality-of-care appraisal in primary care: a quantitative method. Annals of Internal Medicine; 83, 1975, nr. 1, p. 46-52
- SIMPSON, D.E., E.C. RICH, K.A. DALGAARD, D. GJERDINGEN, T.W. CROWSON, D.K. O'BRIEN, P.E. JOHNSON. The diagnostic process in primary care: a comparison of general internists and family physicians. Social Science and Medicine; 25, 1987, nr. 7, p. 861-866
- SINGH, B.M., M.R. HOLLAND, P.A. THORN. Metabolic control of diabetes in general practice clinics: comparison with a hospital clinic. British Medical Journal; 289, 1984, p. 726-728
- SLUIJS, E.M., J.P. DOPHEIDE, J. VAN DER ZEE. Overzichtsstudie onderzoek eerstelijns: stand van het wetenschappelijk onderzoek in en over de eerstelijnsgezondheidszorg en haar raakvlakken. Utrecht: NIVEL, 1985
- SMITH, S.R. Application of the tracer technique in studying quality of care. The Journal of Family Practice; 4, 1977, nr. 3, p. 505-510
- SMITS, A., H. MOKKINK, R. GROL, J. VAN EIJK, M. BEEK, J. MESKER-NIESTEN, P. MESKER. Spreekuurgedrag van huisartsen: de samenhang tussen de medisch-technische en de psychosociale aspecten. Huisarts en Wetenschap; 28, 1985, nr. 4, p. 121-125
- STEWART, M.A. What is a successful doctor: patient interview?: a

- study of interactions and outcomes. *Social Science and Medicine*; 19, 1984, nr. 2, p. 167-175
- STURMANS, F. *Epidemiologie, theorie, methoden en toepassingen*. Nijmegen: Dekker en Van de Vegt, 1982
- TAYLOR, M.P. Epilepsy in a Doncaster practice: audit and change over eight years. *Journal of the Royal College of General Practitioners*; 37, 1987, nr. 296, p. 116-119
- THOMASSEN, J.F.M. *Consultanalyse: communicatieve en kwalitatieve aspecten van hulpverleners in de huisartspraktijk*. Dissertatie. Amsterdam: VU-uitgeverij, 1986
- THOMPSON, H.C., C.E. OSBORNE. Office records in the evaluation of quality of care. *Medical Care*; 14, 1976, nr. 4, p. 294-314
- THOMPSON, M.S., R.H. PALMER, J.K. ROTHROCK, R. STRAIN, L.H. BRACKMAN, E.A. WRIGHT. Resource requirements for evaluating ambulatory health care. *American Journal of Public Health*; 74, 1984, nr. 11, p. 1244-1248
- TIELENS, V., M. BEEK, R. KOOPMANS, Y. VAN TRIER, H. MOKKINK. Behandelt de huisarts vrouwen anders dan mannen? *Huisarts en Wetenschap*, 29, 1986, nr. 7, p. 204
- VUORI, H.V. *Quality assurance of health services: concepts and methodology*. Copenhagen: WHO, 1982, 127 p.
- VUORI, H.V. Patients satisfaction - an attribute or indicator of the quality of care? *Quality Review Bulletin*; 1987, nr. 3, p. 106-108
- WARE, J.E., A. DAVIES-AVERY, A. STEWART. The measurement and meaning of patient satisfaction. *Health and Medical Care Services Review*; 1, 1978, p. 2-14
- WEEL, C. VAN, P.A.M. VAN ZELST. Diabetes mellitus in een huisartspraktijk: II een vervolgonderzoek. *Huisarts en Wetenschap*; 26, 1983, p. 214-217
- Werkgroep toetsing van het Nederlands Huisartsen Genootschap. Toetsing in de huisartsgeneeskunde. *Huisarts en Wetenschap*; 23, 1980, p. 127-140
- WILLIAMSON, J.W. Evaluating quality of patient care: a strategy relating outcome and process assessment. *Journal of the American Medical Association*; 218, 1971, nr. 4, p. 564-569
- WILLIAMSON, J.W. *Assessing and improving health care outcomes: the health accounting approach to quality assurance*. Cambridge: Massachusetts, Ballinger Publishing Company, 1978
- WILSON, A.D. Cervical cytology: a general practice audit. *Family Practice*; 4, 1987, nr. 1, p. 56-60
- WILSON, J.A., T.S. MURRAY. Audit of the drug treatment of Parkin-

- son's disease in general practice. Journal of the Royal College of General Practitioners; 35, 1985, p. 276-278
- WOODWARD, C.A., C.A. MC CONVEY, V. NEUFELD, G.R. NORMAN, A. WALSH. Measurement of physician performance by standardized patients: refining techniques for undetected entry in physicians' offices. Medical Care; 23, 1985, nr. 8, p. 1019-1027
- WORLD HEALTH ORGANIZATION. Statistical indicators for the planning and evaluation of public health programmes. Fourteenth report of the WHO expert committee on health statistics. WHO Technical Report Series no. 472. Geneva: World Health Organization, 1971
- ZUCKERMAN, H.S., J.A. HUNTLEY, K.J. WATERBROOK. Effectiveness of patient care in a primary care clinic. Medical Care; 18, 1980, nr. 10, p. 1001-1012

BIJLAGEN

Bron v. inf. = bron van informatie.

Betrouwbaarh. = betrouwbaarheid.

- Auteur** : Bass e.a., 1986.
- Doel** : Nagaan welke handelingen van huisartsen effect hebben op het resultaat bij de patiënt met inachtneming van andere factoren met prognostische betekenis.
- Populatie** : 232 patiënten uit 13 huisartspraktijken; 83% (193) konden vervolgd worden.
- Aandoeningen** : Buikklachten, rug- of nekpijn, pijn op de borst, vermoeidheidsklachten, hoofdpijn, oogsymptomen, recitaal bloedverlies.
- Handelingen** : Registratie, anamnese, lichamelijk onderzoek, aanvullend onderzoek, verwijzing, diagnostiek, therapie, follow-up, rekening houden met psychosociale factoren.
- Bron v. inf.** : - Interviews met patiënten na 1 en 3 maand(en);
- statussen.
- Criteria** : Impliciete beoordeling van alle afzonderlijke aspecten + overall beoordeling door twee artsen, consensus/gemiddelde; bovendien als extra variabele: waarschijnlijk dat de klacht na 1 maand over zou zijn.
- Resultaten** : - Aantal technische aspecten van zorg **niet** gerelateerd aan resultaat na 1 maand;
- vroeg verdwijnen van klachten gerelateerd aan variabelen betreffende de patiënt, de klacht en ideeën van de patiënt over de arts-patiënt-relatie;
- laat verdwijnen van de klachten gerelateerd aan patiëntgebonden variabelen, klachtgebonden variabelen en registratie van belang psychosociale factoren door arts.
- Conclusie** : Geen verband aantoonbaar tussen verdwijnen van klachten en procescriteria (anamnese, onderzoek, voorschrijven, etc.).
- Betrouwbaarh.** : Geen test uitgevoerd; alle beoordelingen waren 'blind'; hoge mate van overeenstemming gerapporteerd tussen beoordelaars statussen.
- Validiteit** : - Beoordeling statussen lijkt valide;
- patiënteninterviews na 1 maand: positief oordeel over arts-patiënt-relatie, voorafgaand of gevolg van vroeg verdwijnen symptomen?

- Auteur** : Bennett e.a., 1985.
- Doel** : Vergelijking arts-assistenten in traditionele en 'primary care'-opleiding; evaluatie van een nieuw onderwijsprogramma.
- Populatie** : Alle arts-assistenten in het 2^e jaar van deze 2 opleidingen in 4 geaffilieerde klinieken. N = 57 (1040 patiëntencontacten).
- Aandoeningen/ Handelingen** : Screening op colorectale maligniteit; behandeling van hypertensie; voorschrijven van benzodiazepines; behandeling van chronische longaandoeningen.
- Bron van inf.** : Patiëntenstatus.
- Criteria** : - Van te voren vastgesteld door deskundigen; er moest minimaal door 75% van de artsen mee ingestemd zijn;
 - voor ieder criterium aanvaardbaar alternatief;
 - beoordeling door niet-medici; alle statussen die niet voldeden, werden nogmaals beoordeeld door een deskundige op grond van impliciete criteria.
- Resultaten** : - Het enige statistisch significante verschil tussen de primary care groep en de traditionele groep was dat de eersten bij 86% van de patiënten bij wie er een indicatie voor bestond rectaal toucher verrichtten en resultaten van de faeces hemocult test registreerden tegen de laatsten bij 77%. Bij beoordeling na 1 jaar was dit verschil verdwenen.
 - behalve bij het voorschrijven van benzodiazepines en vaccinatie werden door beide groepen hoge niveaus gehaald van voldoen aan de criteria.
- Conclusie** : Met deze meetmethode is er geen significant verschil te vinden tussen arts-assistenten die voor de eerstelijns worden opgeleid en arts-assistenten in de traditionele opleiding.
- Betrouwbaarh.** : Geen test van betrouwbaarheid beschreven.
- Validiteit** : Aangegeven wordt dat het tekortschieten van de meetmethode een onderscheid te maken tussen de 2 groepen te maken kan hebben met te grote flexibiliteit van de methode. De mogelijkheid dat evaluatie op andere gebieden van het handelen van artsen wel onderscheid zou kunnen maken tussen artsen die wel en niet speciaal voor de eerstelijns worden opgeleid, wordt geopperd.

- Auteur** : Borgiel e.a., 1985.
- Doel** : Ontwikkelen en testen van methoden om kwaliteit te meten bij Canadese huisartsen.
- Populatie** : 10 geselecteerde huisartsen, per praktijk steekproef van 40 patiëntenkaarten, per praktijk 102 patiënten.
- Aandoeningen** : 180 diagnostische problemen.
- Handelingen** : Algemeen: bijgehouden patiëntenkaarten, preventie, rationeel voorschrijven.
Per diagnostisch probleem: 11 dimensies van het hulpverleningsproces.
- Bron van inf.** : 1. Vragenlijst huisartsen;
2. samenvattingen patiëntenkaarten;
3. vragenlijst patiënten.
- Criteria** : Expliciet, gebaseerd op consensus.
- Resultaten** : Grote hoeveelheid gegevens.
- Conclusie** : Door beoordeelden en onderzoekers worden de methoden positief beoordeeld, w.b. betrouwbaarheid, validiteit en aanvaardbaarheid.
- Betrouwbaarh.** : Overeenkomst tussen beoordelaars varieert van 64,5% (dimensie 'education') tot 91,2 (dimensie 'hazardous').
- Validiteit** : Wordt door beoordeelde artsen positief beoordeeld; criteria zijn via uitgebreide procedure vastgesteld (maakt dus ook valide indruk).

- Auteur** : Brouwer e.a., 1983.
- Doel** : Evaluatie doelmatigheid aanvragen bij diagnostisch centrum Maastricht.
- Populatie** : Huisartsen in Zuid-Limburg, die gebruik maakten van de diensten van het diagnostisch centrum Maastricht (≥ 1 verwijzing naar DCM/week). N=67.
- Aandoeningen** : Indicaties voor:
- radiodiagnostiek thorax;
 - radiodiagnostiek maag-duodenum;
 - radiodiagnostiek lumbosacrale wervelkolom;
 - mammografie;
 - klinische chemie leverfunctie;
 - klinische chemie schildklier;
 - electrocardiografie.
- Handelingen** : Aanvragen aanvullende diagnostiek.
- Bron van inf.** : Aanvraagformulier.
- Criteria** : - Expliciete werkafspraken door deskundigen, in samenspraak met vertegenwoordigers onderzochte groep;
- Resultaten** : - de gegevensvermelding was na de werkafspraken voor 3 indicaties niet veranderd en voor de overige 4 verbeterd;
- de indicatiestelling was voor de mammografie ruimer gesteld en voor de klinisch-chemische aanvragen en electrocardiografie verbeterd.
- Conclusie** : De DC-activiteiten blijken binnen bepaalde sectoren doelmatig gebruik van diagnostische faciliteiten te hebben bevorderd.
- Betrouwbaarh.** : Geen betrouwbaarheidstest beschreven.
- Validiteit** : Geen validiteitstest beschreven.

- Auteur** : McClure, 1985.
- Doel** : De mogelijkheid om het handelen van family physicians en general physicians te evalueren met behulp van getrainde patiënten wordt onderzocht.
- Populatie** : 26 family physicians van de medische faculteit van de universiteit van Arizona worden geconfronteerd met 5 verschillende gestandaardiseerde patiënten.
- Aandoeningen** : Ongecompliceerde rheumatische aandoeningen.
- Handelingen** : Anamnese (21 verschillende aspecten); lichamelijk onderzoek (7 aspecten); concluderende samenvatting; voorlichting, terugbestellen; therapie en verwijzen.
- Bron van inf.** : 'Checklist', die achteraf wordt ingevuld door een getrainde patiënt; gestructureerde vragenlijsten afgenomen bij de family physicians onder studie.
- Criteria** : Expliciete beoordeling door middel van het 'turven' van handelingen. De checklist is gezamenlijk opgesteld door een family physician en een reumatoloog.
- Resultaten** : Aan de psychosociale aspecten van de ziekte werd door de artsen weinig of geen aandacht besteed. Wel werd vrijwel altijd naar de hoofdklacht en het begin van de klachten gevraagd. Driekwart van de artsen kwam tot een juiste diagnose en vrijwel allen ontwikkelden een adequaat behandelplan.
- Conclusie** : De getrainde standaardpatiënt blijkt een goed instrument om zorgverlening te meten. Van de onderzochte artsen bleek 88% niet gemerkt te hebben een 'standaard'-patiënt geholpen te hebben (blindering). Ook blijkt de 'standaard'-patiënt in 88% van de gevallen een volledig juist beeld van de zorgverlening te hebben gegeven (betrouwbaarheid). Vermeldenswaard is verder dat 'slechts' 19% van de artsen deze methode een inbreuk vond op de privacy.
- Betrouwbaarh.** : Door training en video-evaluatie van het invullen van de checklist wordt toegewerkt naar een intrabeoordelingsbetrouwbaarheid van tenminste 85.
- Validiteit** : Door de welhaast perfecte nabootsing van een normaal arts-patiënt consult met blinding van de arts is deze methode van een goede validiteit. Negatief punt is, dat het consulten van 'nieuwe' patiënten betreft en de vaste relatie arts-patiënt ontbreekt.

- Auteur** : Crebolder en Van Eijndhoven, 1986.
- Doel** : Beoordeling van efficiëntie, compliantie en effectiviteit met betrekking tot zorg voor type-II diabeten na invoering van een speciaal programma hiervoor in een gezondheidscentrum.
- Populatie** : 177 patiënten met diabetes mellitus type II; response: 126.
- Aandoeningen** : Diabetes Mellitus type II.
- Handelingen** : Controle, begeleiding.
- Bron van inf.** : Statussen, interview met 14 patiënten (satisfactie);
- Criteria** : - Impliciet oordeel eigen huisarts (compliantie);
 - impliciet oordeel patiënten (satisfactie);
 - effectiviteit: niet van tevoren vastgesteld wat wel en niet acceptabel is.
- Resultaten** : - 90% van door de huisarts gecontroleerde patiënten komt regelmatig;
 - oordeel patiënten overwegend positief;
 - compliantie lijkt samen te hangen met aantal huisgenoten en kwaliteit gezinsrelatie; niet met bijvoorbeeld leeftijd, geslacht e.a.;
 - effectiviteit: 5-10% slecht ingesteld volgens HbA1c-percentages;
 circa 25% slecht ingesteld volgens nuchter bloedsuikergehalte en quetelet-index.
- Conclusie** : Adequate begeleiding van type II diabeten door de huisarts is zeer wel mogelijk.
- Betrouwbaarh.** : Geen commentaar auteurs.
- Validiteit** : Beoordeling van compliantie en kwaliteit gezinsrelatie door huisarts lijkt niet erg valide; validiteit HbA1c-percentages volgens auteurs hoog.

- Auteur** : Edwards e.a., 1985.
- Doel** : Beoordeling van het beleid bij buikpijnklachten in een huisartsenpraktijk.
- Populatie** : 150 patiënten met buikpijn als hoofdklacht.
- Aandoeningen** : Buik.
- Handelingen** : Anamnese, onderzoek, diagnostiek, beleid.
- Bron van inf.** : - Verslaglegging speciaal voor onderzoek;
- statussen.
- Criteria** : Niet vastgesteld wat wel/niet acceptabel is.
- Resultaten** : Voornamelijk beschrijvend;
- in 21 gevallen kon bij het eerste bezoek geen diagnose gesteld worden;
- achteraf (na een half jaar) bleken minstens 11 van de 150 diagnoses incorrect;
- negen spoedverwijzingen: 8 opgenomen, 3 spoedoperaties (1 ten onrechte) en 1 patiënt overleden;
- van 133 patiënten die niet verwezen waren, waren 125 na een half jaar volledig genezen.
- Conclusie** : De huisarts kan het merendeel van de patiënten met buikpijnklachten zelf behandelen.
- Betrouwbaarh.** : Geen commentaar auteurs.
- Validiteit** : Beoordeling door huisartsen zelf; lijkt valide.

- Auteur** : Fleming en Lawrence, 1983.
- Doel** : Beoordelen van preventieve maatregelen in 29 huisartspraktijken en vergelijken met eerdere beoordeling 2 1/2 jaar eerder.
- Populatie** : 29 praktijken (niet aselekt); in elke praktijk een aselechte steekproef van 200-250 statussen.
- Aandoeningen** : Cervixcarcinoom, rubella, aandoeningen waarvoor algemeen wordt ingeënt, bloeddruk, rookgewoonten.
- Handelingen** : Preventie; registratie van uitstrijkjes, immunisatiestatus, bloeddruk en rookgewoonten.
- Bron van inf.** : Statussen.
- Criteria** : Niet vastgelegd wat wel/niet acceptabel is; vergelijking tussen opeenvolgende analyses.
- Resultaten** : - Verbetering gemiddeld over alle praktijken:
 * cervixcytologie: 56-64%;
 * rubella immunisatie: 28-40%;
 * polio-immunisatie: 15-21%;
 * primaire immunisatie: 68-78%;
 * geregistreerde bloeddruk: 53-61%;
 * geregistreerde rookgewoonten: 22-30%
 (alle significant);
 - verbetering in alle praktijken opgetreden, het sterkst in praktijken die eerst onder het gemiddelde zaten.
- Conclusie** : Preventief handelen van deze artsen verbeterd.
- Betrouwbaarh.** : Geselecteerde groep artsen (eerste onderzoek bij 38 artsen die meededen aan postacademische cursus); geen controlegroep; betrouwbaarheid status-informatie onbekend.
- Validiteit** : Vooral registratie beoordeeld.

- Auteur** : Fletcher e.a., 1984.
- Doel** : Meten van continuïteit en coördinatie als maat voor integratie van hulp door verschillende hulpverleners binnen een General Medicine Practice van een opleidingsziekenhuis.
- Populatie** : Aselecte steekproef van 197 patiënten.
- Aandoeningen/ Handelingen** : Niet van toepassing.
- Bron van inf.** : Patiëntenstatus, patiënteninterview.
- Criteria** : Niet vastgelegd wat wel/niet acceptabel is; expliciete definities voor continuïteit en coördinatie.
- Resultaten** : - Continuïteit (percentage bezoeken aan 'eigen arts'): overall 60%;
 - coördinatie (percentage bezoeken aan andere artsen waarvan vaststond dat de eigen arts ervan af wist): overall 35,2%;
 - grote variatie in resultaten, afhankelijk van in welk deel van de praktijk het contact met de andere arts plaatsvond.
- Conclusie** : Integratie van hulp door verschillende hulpverleners is in bepaalde delen van de onderzochte praktijk hoog; deze methode en definities blijken goed toepasbaar.
- Betrouwbaarh.** : - Geen test beschreven;
 - 'artsen schrijven niet alles in de status'.
- Validiteit** : Geen rekening gehouden met effect van (dis)continuïteit; belangrijk: wat is het effect voor de patiëntenzorg?

- Auteur** : Gerbert en Hargreaves, 1986.
- Doel** : Vier verschillende methoden van informatie verzamelen over gedrag van artsen beoordelen en vergelijken wat betreft betrouwbaarheid, validiteit, kosten en aanvaardbaarheid.
- Populatie** : 63 'primary care physicians' (vrijwilligers) en 214 door hen geselecteerde volwassen CARA-patiënten.
- Aandoeningen** : Chronisch obstructieve longaandoeningen.
- Handelingen** : Anamnese, lichamelijk onderzoek, aanvragen aanvullend onderzoek, behandeling en patiënteneducatie.
- Bron van inf.** : 1. Artseninterviews;
2. patiënteninterviews;
3. status;
4. video-opname.
- Criteria** : Expliciete criteria ('criteria map'), beoordeling door onderzoekers.
- Resultaten** : Betrouwbaarheid (interrater) hoog ($> 0,80$) bij alle vier methoden.
Inhoudsvaliditeit over het algemeen goed, bij andere twee methoden matig.
Convergente validiteit over het algemeen goed; bij combinatie patiënteninterview-status het laagste.
Kosten niet erg uiteenlopend, laagste voor status.
Aanvaardbaarheid goed.
- Conclusie** : Om optimaal resultaat te krijgen, kunnen het beste twee of meer methoden tegelijkertijd gebruikt worden, waaronder in ieder geval een interviewmethode.
- Betrouwbaarh.** : Zie resultaten.
- Validiteit** : Zie resultaten.

Auteur : Gerritsma en Smal, 1982.

Doel : Verschillen nagaan tussen werkwijze van huisartsen en internisten.

Populatie : 16 huisartsen en 16 internisten; niet aselekt gekozen.

Aandoeningen/ Handelingen : Volledig medisch handelen bij:
 - Ulcus duodeni met/zonder bloeding;
 - hoofdpijn, matige hypertensie met/zonder fenacetinegebruik.

Bron van inf. : Papieren patiënten.

Criteria : Geen kwaliteitscriteria; verschillen tussen groepen werden nagegaan; hierover 13 hypothesen opgesteld.

Resultaten : Internisten gaan uitvoeriger te werk, met name wat betreft somatische aspecten; tussen huisartsen onderling grotere verschillen in werkwijze dan tussen internisten.

Conclusie : Resultaten geven inzicht in de wijze van medisch denken en handelen.

Betrouwbaarh. : Bij simulaties moeilijk te bepalen.

Validiteit : Uitgebreid onderzocht aan de hand van vooraf opgestelde schattingen; resultaten positief.

- Auteur** : Mc Ghan e.a., 1983.
- Doel** : Vergelijking van de kwaliteit van het voorschrijven van antihypertensiva aan ambulante patiënten door apothekers en artsen in een HMO.
- Populatie** : 2 apothekers en 3 artsen; in elk van de twee groepen een aselechte steekproef van circa 150 voorschriften voor antihypertensiva.
- Aandoeningen** : Hypertensie.
- Handelingen** : Beoordeling werkzaamheid, bijwerkingen en therapietrouw, laboratoriumonderzoek, voorschrijven en verwijzen.
- Bron van inf.** : Gecomputeriseerde gegevens van statussen ('blind').
- Criteria** : Expliciete criteria voor voorschrijven, gebaseerd op rapport van de Joint National Committee on Detection, Evaluation and Treatment of High Blood Pressure (1980) en beoordelingsschaal (impliciet), beide toegepast door een klinische farmacoloog en een ervaren klinische apotheker (onafhankelijk van elkaar).
- Resultaten** : - Geen significant verschil tussen beide groepen patiënten;
 - apothekers waren significant beter ($p < 0,01$) in het kiezen van het juiste medicijn volgens de expliciete criteria;
 - volgens de impliciete criteria geen significant verschil;
 - overall scoorden de apothekers significant beter ($p < 0,05$);
 - in een steekproef van de patiënten bleek na het voorschrijven de diastolische druk bij patiënten uit de apothekergroep iets lager (niet significant).
- Conclusie** : De apothekers schreven zeker zo goed voor als de artsen (zowel op proces- als resultaatniveau).
- Betrouwbaarh.** : Kleine groepen, niet aselekt; betrouwbaarheid computerinformatie niet bepaald; interrater reliability coëfficiënt 0,65 ($p < 0,001$); beoordeling 'blind'.
- Validiteit** : Outcome-resultaten wijzen in dezelfde richting als proces.

Auteur : Given e.a., 1981.

Doel : Beoordeling van 'family practice residents' op basis van een gecomputeriseerd informatie-systeem.

Populatie : 25 family practice residents.

Aandoeningen : Alle voorkomende diagnoses.

Handelingen : - Diagnose stellen;
- aanvragen lab/Rö-onderzoek;
- voorschrijven;
- patiënt laten terugkomen.

Bron van inf. : - Computersysteem met informatie over ieder arts-patiënt;
- statussen.

Criteria : 50% of meer onder of 100% of meer boven het gemiddelde van het hele centrum.

Resultaten : Per 'resident' indruk van zwakke/sterke kanten.

Conclusie : Het systeem is zeer bruikbaar als snelle screening van residents.

Betrouwbaarh. : Niet duidelijk hoe betrouwbaar de informatie is die wordt aangeleverd door 'residents' voor het informatiesysteem.

Validiteit : 'Residents' die weinig van het gemiddelde afwijken, maar toch inadequate zorg verlenen, worden gemist.

Auteur : Grace en Armstrong, 1987.

Doel : Oordeel over verwijsgedrag van huisartsen van patiënten, specialisten en huisartsen vergelijken.

Populatie : 306 nieuwe verwijzingen van 114 huisartsen.

Aandoeningen : Niet van toepassing.

Handelingen : Verwijzen.

Bron van inf. : Vragenlijsten.

Criteria : Impliciet oordeel specialist, huisarts en patiënt.

Resultaten : - Specialisten zijn het meest kritisch (kritiek bij helft van de gevallen);
- patiënten zijn kritisch omtrent gedrag huisarts vóór verwijzing, minder omtrent mogelijkheden van huisartsen zelf;
- huisartsen minst kritisch (kritiek bij éénderde van de gevallen);
- weinig overlap.

Conclusie : Weinig overeenstemming tussen huisartsen, patiënten en specialisten over wat een huisarts kan doen en naar wie hij dient te verwijzen.

Betrouwbaarh. : - Deels elders beschreven;
- patiëntengroep vertekend door non-reponse.

Validiteit : Subjectieve oordelen.

- Auteur** : Grol e.a., 1985.
- Doel** : Verband nagaan tussen taakopvatting en kwaliteit van het handelen van huisartsen.
- Populatie** : 57 niet-aselecte huisartsen.
- Aandoeningen** : 24 protocolklachten.
- Handelingen** : Attitudes: taakafbakening ten opzichte van specialisten somatisch, competentie gevoel somatisch; taakafbakening ten opzichte van specialisten psychosociaal, competentiegevoel psychosociaal, medisch handelen, niet-somatisch handelen.
- Bron van inf.** : Vragenlijsten, ziekenfondsgegevens, observaties en audio-opnames.
- Criteria** : Protocollen voor medisch handelen, impliciet oordeel over niet-somatisch handelen, veel/weinig verwijzen op basis van factoranalyse.
- Resultaten** : Samenhang tussen aspecten van taakopvatting enerzijds en verwijsgedrag en kwaliteit van het handelen (voldoende diagnostiek en therapie en overbodige handelingen) anderzijds.
- Conclusie** : Opvattingen w.b. taakopvatting en competentie hebben belangrijke consequenties voor de geboden zorg.
- Betrouwbaarh.** : Medisch handelen hoge interrater reliability, psychosociaal handelen minder hoog, betrouwbaarheid attitudes: alpha 0,63-0,79.
- Validiteit** : Gevonden verband pleit voor validiteit van verschillende meetmethoden.

- Auteur** : Grol e.a., 1986a.
- Doel** : Nagaan of een programma van onderlinge toetsing effect heeft op de handelwijze van huisartsen.
- Populatie** : 43 niet-aselecte huisartsen; per huisarts 15-20 voormetingen (1/2-1 jaar vóór project), 13-15 name-tingen (1/2-1 jaar na project).
- Aandoeningen** : 24 protocolklachten.
- Handelingen** : - Medisch-technisch handelen;
- niet-medisch handelen;
- Bron van inf.** : Observaties en audio-opnames.
- Criteria** : Deels expliciet (medisch-technisch: obligate en overbodige handelingen), deels impliciet (niet-medisch handelen: gericht en systematisch werken, hanteren van de huisarts-patiënt relatie en psychosociaal handelen).
- Resultaten** : - Meer handelen in overeenstemming met de normen;
- resultaten sterker verbeterd bij huisartsen die eerst lager scoorden.
- Conclusie** : Nader onderzoek vereist.
- Betrouwbaarh.** : - Medisch-technisch handelen: overeenstemmingsper-centage 0,71-0,75;
- Niet-medisch handelen: Kendall's tau 0,30-0,50 bij de voor- en 0,40-0,60 bij de nameting;
- geen aselecte steekproef;
- geen controlegroep.
- Validiteit** : Normen en criteria vastgesteld na uitgebreide dis-cussies; verband met outcome ligt überhaupt vrijwel nooit vast.

Auteur : Grol e.a., 1986b.

Doel : Effect nagaan van programma van onderlinge toetsing op voorschrijf- en verwijsgedrag van huisartsen.

Populatie : Experimentele groep en controlegroep (geen aselechte toewijzing) totaal 115 huisartsen; cijfers over twee jaren vóór toetsing en twee jaren tijdens toetsingsprogramma.

Aandoeningen : Niet van toepassing.

Handelingen : - Voorschrijven: gericht en aspecifiek voorschrijfgedrag;
- verwijzen naar specialismen samen.

Bron van inf. : Ziekenfondsgegevens.

Criteria : Niet vastgelegd wat wel en niet acceptabel is.

Resultaten : - Daling in het totale voorschrijven het sterkst in de experimentele groep; aspecifiek voorschrijfgedrag daalt, gericht voorschrijven stijgt;
- verwijzen: geen significante verschillen.

Conclusie : Deelname aan het toetsingsprogramma heeft wel invloed gehad op het (met name aspecifiek) voorschrijven, maar niet op het verwijzen.

Betrouwbaarh. : - Ziekenfondscijfers beperkt;
- geen aselechte toewijzing experimentele/controlegroep.

Validiteit : Elders beschreven.

- Auteur** : Grol, 1987.
- Doel** : Onder andere effect nagaan van een programma voor onderlinge toetsing bij huisartsen op het handelen van huisartsen en hun attitude.
- Populatie** : 43 deelnemende huisartsen (geen aselechte groep) (voor onderdeel 'voorschrijven en verwijzen': 28 deelnemende en 56 niet-deelnemende huisartsen).
- Aandoeningen** : Alle voorkomende aandoeningen: 'niet-medisch handelen'; 'medisch handelen': protocollen voor 24 veel voorkomende nieuwe klachten en controles.
- Handelingen** : Volgens protocollen: obligate en overbodige handelingen; 'niet-medisch handelen': gericht en systematisch handelen; hanteren van huisarts-patiënt-relatie; psychosociaal handelen.
- Bron van inf.** : Directe observaties, audiobanden, vragenlijsten over attitude, ziekenfondsgegevens over voorschrijven en verwijzen.
- Criteria** : Expliciete criteria voor 'medisch handelen' en 'niet-medisch handelen'.
- Resultaten** : - 'Medisch handelen':
 * aantal verrichte obligate handelingen nam toe van 45% naar 51%;
 * aantal verrichte overbodige handelingen nam toe van gemiddeld 1,0 naar 1,4 per consult.
 - 'Niet-medisch handelen': duidelijke verbetering;
 * attitude: weinig verandering;
 * voorschrijven: vooral specifiek voorschrijven nam af.
- Conclusie** : Het is mogelijk het handelen van huisartsen te veranderen met behulp van onderlinge toetsing.
- Betrouwbaarh.** : Interbeoordelaarbetrouwbaarheid observatie-instrumenten;
 'medisch handelen': kappa = 0,71-0,75;
 'niet-medisch handelen': Kendall's tau = 0,30-0,60.
- Validiteit** : Inhoudsvaliditeit van protocollen en criteria goed. Begripsvaliditeit: sterke onderlinge samenhang tussen scores op verschillende protocollen; samenhang tussen scores op beoordelingsinstrumenten en attitude vragenlijsten en voorschrijf- en verwijzroutines; instrumenten discrimineren tussen onervaren en ervaren huisartsen.

- Auteur** : Grol e.a., 1988.
- Doel** : Effect nagaan van een programma van onderlinge toetsing op de kwaliteit van het handelen van huisartsen en op hun voorschrijfgedrag.
- Populatie** : 43 niet-aselecte huisartsen (experimentele groep): voor- en nameting kwaliteit;
Voorschrijven: experimentele groep 19 niet-apotheekhoudende huisartsen; controlegroep 45 niet-apotheekhoudende huisartsen.
- Aandoeningen** : 24 protocolklachten.
- Handelingen** : Medisch-technisch handelen, niet-medisch handelen, voorschrijven.
- Bron van inf.** : - Medisch-technisch handelen en niet-medisch handelen: observatie en audio-opnames;
- voorschrijven: ziekenfondsgegevens.
- Criteria** : Expliciete protocollen voor medisch-technisch handelen, impliciete beoordeling niet-medisch handelen; voor voorschrijven geen vastlegging van wat wel/niet acceptabel is.
- Resultaten** : - 74% van de deelnemers werkten na het programma meer volgens de expliciete criteria, vrijwel alle deelnemers verbeterden niet-medisch handelen, zoals brede oriëntatie, verduidelijken van vragen en verwachtingen van de patiënt;
- de huisartsen met lage beginscore verbeterden het meest;
- in de experimentele groep gemiddeld een half voorschrift minder per patiënt tijdens het programma, in de controlegroep gemiddeld een half voorschrift meer.
- Conclusie** : Door het programma zijn huisartsen meer volgens de criteria gaan werken.
- Betrouwbaarh.** : Medisch-technisch handelen: interreliability voormeting 0,30-0,50, nameting 0,40-0,60.
Geen aselecte toewijzing experimentele/controlegroep.
- Validiteit** : Criteria vastgesteld na uitgebreide procedure.

- Auteur** : Haaijer-Ruskamp e.a., 1983.
- Doel** : Nagaan in hoeverre huisartsen verschillen in voorschrijfgedrag bij dezelfde klacht.
- Populatie** : Niet-apotheekhoudende huisartsen in Twente. N = 118 (response rate 90%).
- Aandoeningen** : - Hypertensie;
- slapeloosheid;
- knieletsel;
- onrust bij bejaarden;
- keelpijn;
- gemaskeerde depressie.
- Handelingen** : Voorschrijven.
- Bron van inf.** : Antwoorden op schriftelijke patiëntensimulatie.
- Criteria** : Impliciet, beoordeling door 6 deskundigen (consensus).
- Resultaten** : - Grote verschillen tussen huisartsen in geneigdheid om voor te schrijven en in keuze preparaat;
- al dan niet voorschrijven is zowel arts- als klachtgebonden;
- farmacologisch juist voorschrijven is artsgebonden;
- juistheid van voorschrijfduur en dosering geneesmiddelgebonden;
- voorschrijfgedrag bij de schriftelijke simulaties komt overeen met werkelijk gedrag op basis van ziekenfondsgegevens, zowel wat betreft geneigdheid voor te schrijven als wat betreft juistheid van voorschrijven (volgens farmacotherapeutisch kompas).
- Conclusie** : Huisartsen variëren aanzienlijk in hun voorschrijfgedrag, zowel kwalitatief als kwantitatief; implicaties voor opleiding.
- Betrouwbaarh.** : Geen test beschreven.
- Validiteit** : Het intentionele gedrag zoals gemeten wordt met behulp van de schriftelijke simulaties, blijkt overeen te komen met werkelijk voorschrijfgedrag; dit pleit voor de validiteit van de meetmethode.

- Auteur** : **Hawton e.a., 1987.**
- Doel** : Nagaan of poliklinische nazorg voor patiënten die zelfmoordpogingen hebben gedaan zinvol is of bepaalde groepen patiënten er meer baat bij hebben dan andere.
- Populatie** : 80 patiënten die overdosis medicijnen hadden genomen en voor wie psychiatrische opname niet nodig was, aselekt verdeeld in 2 groepen.
- Aandoeningen** : Zelfmoordpoging met medicamenten.
- Handelingen** : Begeleiding.
- Bron van inf.** : Uitkomsten:
- van getrainde beoordelaar;
- van patiënt:
* Suicidal Intent Scale;
* Risk of Repetition Scale;
* Social Adjustment Scale;
* Beck depression Inventory;
* General Health Questionnaire;
- van huisarts.
- Criteria** : Statistisch verschil in outcome tussen 2 groepen waarvan er één poliklinische begeleiding kreeg en de andere door de huisarts begeleid werd.
- Resultaten** : Geen significant verschil in outcome tussen beide groepen; bij vrouwelijke patiënten en patiënten met relatieproblemen blijkt poliklinische begeleiding beter resultaat te hebben dan begeleiding door de huisarts.
- Conclusie** : - Niet alle patiënten hebben na een zelfmoordpoging poliklinische hulp nodig;
- compliance moet verbeterd worden;
- patiënten met relatieproblemen moeten bij twijfel aan eigen therapeutische mogelijkheden door de huisarts worden doorverwezen.
- Betrouwbaarh.** : De interrater reliability coëfficiënten waren hoog voor de beoordelaars: 0,83, 0,83 en 0,86; $p < 0,001$. De patiëntenvragenlijsten werden zo overgenomen zonder vermelding van de betrouwbaarheid.
- Validiteit** : Over de validiteit van de vragenlijsten wordt niet gerapporteerd, evenmin over die van de externe beoordeling.

- Auteur** : Hayes en Harries, 1984.
- Doel** : Nagaan of patiënten met ongecompliceerde diabetes mellitus type II even goed in de huisartspraktijk begeleid kunnen worden als in het ziekenhuis.
- Populatie** : 200 patiënten 40-80 jaar uit ziekenhuispopulatie met ongecompliceerde diabetes, at random verdeeld in 2 groepen.
- Aandoeningen** : Ongecompliceerde diabetes mellitus type II.
- Handelingen** : Begeleiding.
- Bron van inf.** : Patiënteninterview + -onderzoek.
- Criteria** : Niet vastgelegd wat wel/niet acceptabel is; vergelijkend onderzoek.
- Resultaten** : - 13,6% van de patiënten in de huisartsengroep werd minstens één maal per jaar gezien, bij 4,8% werd minstens één maal per jaar bloedglucose bepaald; in de ziekenhuisgroep was dit 100%;
 - HbA1 na 5 jaar: 10,4% (sd. 1,73) in huisartsengroep, 9,5% (sd. 1,77) in ziekenhuisgroep (t = 2,52, p < 0,02);
 - in de huisartsengroep stierven 18 patiënten, tegenover 6 in de ziekenhuisgroep ($\chi^2 = 5,642$, p 0,02).
- Conclusie** : Routinematige zorg voor ongecompliceerde type II diabetes is in de huisartspraktijk minder bevredigend dan in het ziekenhuis.
- Betrouwbaarh.** : - Randomisatie!
 - betrouwbaarheid gegevensverzameling niet becommentarieerd (met name niet 'blind').
- Validiteit** : Gezien de doodsoorzaken (oversterfte aan myocardinfarct in huisartsengroep) lijkt de sterfte inderdaad samen te hangen met minder controles; helaas geen andere complicaties bepaald (met name retinopathie); hemoglobine A concentratie wordt algemeen beschouwd als indicator voor goede instelling diabetes; helaas niet duidelijk hoe groot de klinische relevantie is van het gevonden verschil.

Auteur : Heineken e.a., 1985.

Doel : Toepasbaarheid nagaan van een kwaliteitsmeting die gebruik maakt van 'negative indexes of health'.

Populatie : 1.147 patiënten van een eerstelijnsgrupspraktijk.

Aandoeningen : Lijst A en lijst B van Rutstein e.a.
(lijst A: aandoeningen waarvan ieder enkel geval nader onderzoek verdient);
(lijst B: aandoeningen waarvan vóórkomen boven bepaald aantal onderzoek nodig maakt).

Handelingen : Primaire en secundaire preventie.

Bron van inf. : - Gecomputeriseerd informatiesysteem;
- statussen.

Criteria : Semi-expliciet.

Resultaten : Geen van de gevallen uit lijst A was te voorkomen.
Van de gevallen uit lijst B was alleen een deel van 'complicaties van hypertensie' te voorkomen; behandeling van hypertensie in jaar, voorafgaand aan optreden van complicaties, voldeed echter.

Conclusie : - Methode bevordert preventie;
- gevoeligheid methode moeilijk te bepalen door klein aantal gevallen.

Betrouwbaarh. : Wordt niet duidelijk.

Validiteit : Zie conclusie.

Auteur	: Hill e.a., 1978.
Doel	: Vergelijken van de effectiviteit van begeleiding van patiënten verdacht van myocardinfarct, door de huisarts en in de ziekenhuissituatie.
Populatie	: Patiënten uit de praktijken van 60 huisartsen, verdacht van myocardinfarct, zonder complicaties (beoordeeld door ambulance-team), aselekt verdeeld in 2 groepen van 132 patiënten.
Handelingen	: Verdenking van myocardinfarct zonder complicaties.
Bron van inf.	: - Gegevens ambulance-team, laboratoriumuitslagen; - beoordeling patiënten na 6 weken.
Criteria	: Niet vastgelegd wat wel/niet acceptabel is; vergelijkend onderzoek.
Resultaten	: - 26 patiënten uit de thuisgroep moesten later toch opgenomen worden (falen van beleid thuis); - mortaliteit na 6 weken niet significant verschillend in beide groepen.
Conclusie	: Het is mogelijk met een ambulance-team, zoals hier toegepast, een groep patiënten verdacht van myocardinfarct te identificeren die even goed door de huisarts als in het ziekenhuis behandeld kunnen worden.
Betrouwbaarh.	: - Randomisatie! - betrouwbaarheid gegevensverzameling na randomisatie niet bekend (blind?).
Validiteit	: Mortaliteit lijkt een zeer valide outcome-variabele bij infarct-patiënten.

Auteur	: Holmes e.a., 1978.
Doel	: Vergelijken van family physicians die wel en niet formeel onderwijs in family practice hebben gevolgd op basis van criteria voor de specifieke aspecten van family practice.
Populatie	: 14 artsen die het family practice-onderwijs hebben gevolgd in een bepaald gebied (RT), gematchte controlegroep (op basis van urbanisatiegraad, solo- of groepspraktijk, aantal jaren sinds beëdiging) (PE).
Aandoeningen	: Niet van toepassing.
Handelingen	: 4 karakteristieken van family practice: continuïteit, coördinatie, veelomvattendheid, gerichtheid op het gezin.
Bron van inf.	: Statussen.
Criteria	: Door middel van gemodificeerde Delphimethode 7 expliciete criteria voor de 4 karakteristieken vastgelegd; niet vastgelegd wat wel/niet acceptabel is.
Resultaten	: - Continuïteit: geen significant verschil; - coördinatie: geen significant verschil; - veelomvattendheid: in minder dan 5% informatie over andere dan hoofdklacht (beide groepen); bij kinderen: RT-artsen meer vastgelegd; - gezinszorg: bij 5 van de 14 RT-artsen geen informatie over gezin; bij de overige 9 paren RT/PE: RT significant beter.
Conclusie	: Bepaalde aspecten van family practice worden op hoog niveau toegepast, andere niet; vrij weinig verschil tussen al dan niet formeel getrainde family physicians.
Betrouwbaarh.	: 2 beoordelaars, niet 'blind', geen interrater reliability bepaald.
Validiteit	: Criteria via uitgebreide procedure vastgesteld; lijken valide, mogelijk niet gevoelig genoeg.

- Auteur** : Hulka e.a., 1976.
- Doel** : Vergelijken van de kwaliteit van zorg voor 4 indicator aandoeningen tussen family physicians en specialisten.
- Populatie** : Aselecte steekproef, response rate 68%, uiteindelijk: 34 family physicians, 11 internisten; 8 kinderartsen en 8 obstetrici.
- Aandoeningen** : Op latere leeftijd begonnen diabetes mellitus, decompensatio cordis, normale zwangerschap, normale zuigeling tot 1 jaar.
- Handelingen** : Instructies, lichamelijk onderzoek, laboratoriumonderzoek, inentingen.
- Bron van inf.** : - Statussen;
- vragenlijsten aan artsen.
- Criteria** : Expliciete minimumcriteria ontwikkeld door ad hoc commissie, na literatuuronderzoek.
- Resultaten** : - Gemiddelde scores (maximaal 1) hoog voor alle 4 aandoeningen (0,78-0,93), grote variatie per aandoening (bijvoorbeeld diabetes 0,26-1,00);
- voor de zwangerschap en zuigelingen zorg bleek het specialisme van de arts een groot deel van de variatie te verklaren, bij de andere 2 aandoeningen zijn noch arts-, noch patiëntvariabelen gecorreleerd aan de scores.
- Conclusie** : De methode dient verder ontwikkeld en gevalideerd te worden.
- Betrouwbaarh.** : Betrouwbaarheid case-abstractors niet bepaald; betrouwbaarheid vragenlijsten lijkt twijfelachtig.
- Validiteit** : Kwaliteit registratie niet te beoordelen; vaststelling criteria lijkt valide, zal verder gevalideerd worden aan outcomes, echter mogelijk niet gevoelig genoeg.

- Auteur** : **Imants, 1979.**
- Doel** : Nagaan waarin huisartsen bij het opsporen en behandelen van bloeddruk afwijken van de richtlijnen en wat daarvan de oorzaken zijn.
- Populatie** : 23 huisartsen die aan andere onderzoeken hadden meegewerkt, 168 huisartsen die aan nascholingscursus deelnamen.
- Aandoeningen** : Hypertensie.
- Handelingen** : Opsporen; bloeddruk meten; sfeer tijdens consult; informatie aan patiënt; beleid.
- Bron van inf.** : Video-opnamen; antwoorden op vragenlijsten; letterlijke verslagen van discussies.
- Criteria** : Deels expliciete criteria van het NHI, deels subjectief oordeel, beoordeling door 2 artsen.
- Resultaten** : In een aantal gevallen weken de artsen duidelijk af van de richtlijnen, bijvoorbeeld slechts in 9% van de gevallen werd in duplo gemeten, in 33% van de gevallen werd een andere manometer gebruikt dan aanbevolen, in geen van 24 gevallen werd zoals aanbevolen de klok van de stethoscoop gebruikt, in alle 82 gevallen daalde de kwikkolom te snel volgens de richtlijnen, bij 23% van de patiënten met verhoogde bloeddruk werd geen mededeling gedaan over de uitkomst van de meting.
- Conclusie** : Aanzienlijke afwijkingen van de richtlijnen; richtlijnen zijn ook matig bekend bij huisartsen die aan nascholing deelnemen.
- Betrouwbaarh.** : Hoewel alle videobanden door 2 artsen beoordeeld zijn, wordt de overeenkomst tussen de 2 beoordelaars niet beschreven. Een groot aantal observaties kon niet gedaan worden door de opstelling. De op video opgenomen huisartsen zijn anderen dan zij die de vragenlijsten beantwoordden; de video-opnamen werden gemaakt voordat de richtlijnen gepubliceerd waren; daardoor zijn de uitkomsten van beide methoden niet goed vergelijkbaar.
- Validiteit** : De video-opnamen zijn voor een ander doel gemaakt; het meten van de bloeddruk werd daardoor dus niet beïnvloed; bij de vragenlijstmethode wordt intentie gemeten, de validiteit van directe observatie is groter.

Auteur : Kane, 1977.

Doel : Vergelijken van 2 soorten procesmaten met resultaatmaten.

Populatie : 2 centra voor 'family practice'; 410 acute episodes.

Aandoeningen : Alle voorkomende aandoeningen, waarbij voor de volgende 7 aandoeningen expliciete criteria gehanteerd werden: otitis media, hypertensie, bronchitis, pharyngitis, tonsillitis, recidiverende urineweginfectie, vaginitis.

Handelingen : - Diagnostiek en therapie (expliciete criteria);
- gegevensverzameling, probleemherkenning, behandeling, follow-up (implementatie-criteria).

Bron van inf. : Statussen.

Criteria : 1. Voor 7 aandoeningen (251 casussen) expliciete criteria, ontwikkeld door voorloper PSRO;
2. Voor alle overige aandoeningen (169 casussen) impliciete beoordeling door artsen verbonden aan de faculteit.

Resultaten : Beide procesbeoordelingen: positief verband tussen proces en resultaat in termen van functionele status.
Verband met tevredenheid niet duidelijk door klein aantal ontevreden patiënten.

Conclusie : Het is mogelijk een verband te vinden tussen proces en resultaat bij bepaalde aandoeningen in ambulante zorg.

Betrouwbaarh. : Hierover geeft de auteur geen commentaar.

Validiteit : Expliciete criteria zouden van tevoren getest zijn; over validiteit geen commentaar; ook geen vergelijking impliciet oordeel met expliciete criteria.
Ook over validiteit resultaatcriteria niets meege-deeld.

- Auteur** : **Kaptein e.a., 1987.**
- Doel** : Nagaan of huisartsen kwalitatief goede zorg leveren aan CARA-patiënten.
- Populatie** : 150 CARA-patiënten, aangemeld door 10 huisartsen, tussen 1 en 55 jaar.
- Aandoeningen** : Asthma.
- Handelingen** : Voorschrijven van medicatie.
- Bron van inf.** : Vragenlijsten aan patiënt.
- Criteria** : Expliciete criteria voor onderbehandeling aan de hand van resultaat bij patiënt in combinatie met voorgeschreven medicatie.
- Resultaten** : 66 patiënten (44%) waren onderbehandeld in relatie tot één of meer van de vier bestudeerde symptomen.
- Conclusie** : Asthma wordt in Nederland onderbehandeld.
- Betrouwbaarh.** : Niet bepaald; wel vermeld dat patiënten vaak neigen tot onderrapportage van symptomen.
- Validiteit** : Discussie omtrent criteria over onderbehandeling mogelijk; symptomen ook afhankelijk van bijvoorbeeld therapietrouw van de patiënten.

- Auteur** : Kekki, 1987.
- Doel** : Evaluatie van de kwaliteit van het handelen van huisartsen.
- Populatie** : 5 huisartsen in een gezondheidscentrum; gestratificeerde steekproef van 2.540 patiënten uit alle patiënten die gedurende één jaar het gezondheidscentrum bezochten.
- Aandoeningen** : Alle klachten waarmee patiënten kwamen; nader onderzoek naar: otitis media en luchtweginfectie bij kinderen 0-4 jaar, keelpijn, luchtweginfecties, bovenbuikspijn, lage rugpijn, hypertensie, vaginitis, urineweginfecties en diabetes mellitus bij patiënten van 25-49 jaar, hypertensie en diabetes mellitus bij patiënten ouder dan 65.
- Handelingen** : Uiteenlopend.
- Bron van inf.** : Patiëntenkaarten.
- Criteria** : Niet duidelijk of er sprake was van expliciete criteria; maakt meer de indruk van 'screening' met achteraf vaststellen of de zorg voldeed.
- Resultaten** : Uiteenlopend; volgens auteur veel voor verbetering vatbaar, maar ook diepergaand onderzoek nodig naar aspecten die bij 'screening' niet bleken te voldoen.
- Conclusie** : Onderzoek naar de inhoud van het handelen van huisartsen is essentieel.
- Betrouwbaarh.** : Niet besproken door auteur; wel aangegeven dat patiëntenkaarten onvoldoende informatie bevatten (deels kwaliteitsaspect, deels betrouwbaarheidsaspect).
- Validiteit** : Wordt nadelig beïnvloed door de onvolledigheid van de informatie.

Auteur : Kemeny e.a., 1984.

Doel : Onderzoeken van inhoudsvaliditeit, meetbaarheid, kosten en aanvaardbaarheid van vier methoden om de kwaliteit van het handelen van artsen te beoordelen.

Populatie : 13 artsen (vrijwilligers), 30 patiënten.

Aandoeningen : Chronisch-obstructieve longaandoeningen.

Handelingen : Geneesmiddelenbeleid, behandeling van bijwerkingen, onderzoek, patiënteneducatie, navragen naar complicaties, behandeling van exacerbaties, behandeling van complicaties.

Bron van inf. : 1. Artseninterviews;
2. patiënteninterview;
3. video-opname;
4. patiëntenstatus.

Criteria : 'Criteria-map' voor CARA.

Resultaten : Interviewmethoden meest volledig (inhoudsvaliditeit), patiëntenstatus minst volledig, kosten hoogste voor video-opnamen.

Conclusie : Geen van deze vier methoden duidelijk superieur.

Betrouwbaarh. : Niet genoemd.

Validiteit : Zie resultaten.

- Auteur** : Kramer, 1984.
- Doel** : Met behulp van de case-control-methode nagaan of kinderartsen beter dan algemene artsen in staat zijn ernstige acute ziekten bij kinderen te herkennen of voor preventie toegankelijke complicaties te voorkomen.
- Populatie** : 103 case-kinderen, met complicaties die voorkomen hadden kunnen worden of niet tijdige diagnose, therapie of verwijzing; 103 controles, gematcht op leeftijd, SEK en soort ziekte, 152 verschillende artsen die door deze kinderen bezocht waren.
- Aandoeningen** : - Gastro-enteritis;
- meningitis/koorts;
- pneumonie/respiratoire aandoening;
- otitis media.
- Handelingen** : - Voorkomen van complicaties;
- tijdige diagnostiek, therapie en/of verwijzing.
- Bron van inf.** : - Interview met ouders;
- status bij aanmelding eerste hulp;
- provinciaal + nationaal artsenregister.
- Criteria** : Expliciete criteria over onderscheid case-controle.
- Resultaten** : Oddsratio, omega (schatting van relatief risico van ongewenste uitkomst na contact met kinderarts) is 1,32 (95% betrouwbaarheidsinterval 0,76-2,29).
- Conclusie** : - Zorg van kinderartsen is niet beter dan die van algemene artsen op de hier onderzochte aandoeningen;
- case-control-methode is goed toepasbaar voor kwaliteitsmeting.
- Betrouwbaarh.** : - Informatie 'blind' verkregen (dus zonder dat specialisme arts bekend was);
- verder geen betrouwbaarheid bepaald.
- Validiteit** : Beoordelen cases/controls op voorkomen complicaties, danwel tijdig diagnosticeren/therapie instellen/verwijzen maakt een valide indruk.
Geen commentaar van auteurs omtrent validiteit.

Auteur : Kurji en Haines, 1984.

Doel : Nagaan hoe de kwaliteit is van de zorg voor hypertensie-patiënten door huisartsen.

Populatie : 900 patiënten tussen 30 en 65 jaar, aselechte steekproef uit de populatie van 18 huisartsen uit het Brent Health District (response rate 78% uit een aselechte steekproef van 28 artsen).

Handelingen : Case-finding, registratie, follow-up, behandeling.

Bron van inf. : - Uittreksel uit 'practice notes' en correspondentie;
- gestandaardiseerd interview met arts over handwijze waarbij registratie en beleid ten aanzien van hypertensie.

Criteria : Expliciete beschrijving van registratie, vervolgreregistratie en behandeling, zonder duidelijke vaststelling van wat wel en niet acceptabel is.

Resultaten : Beschrijvend.

Betrouwbaarh. : Geen commentaar auteurs.

Validiteit : Geen commentaar auteurs.

Auteur : Kuyvenhoven e.a., 1986.

Doel : Beoordelen van validiteit en betrouwbaarheid van meetinstrument om verschillen tussen huisartsen bij beleid bij patiënten met vage klachten te beoordelen.

Populatie : 19 huisartsen (geselecteerde groep).

Aandoeningen : Vage klachten.

Handelingen : Handelwijze bij diagnostiek en beleid.

Bron van inf. : Antwoorden op schriftelijk simulatie.

Criteria : - Impliciet, beoordeeld door huisartsen-experts.
 - 'voldoende aandacht aan lichamelijke oorzaken/aspecten?';
 - 'kans op onnodige schade?'.

Resultaten : - Subjectief overeenstemming met handelen in werkelijkheid;
 - objectief overeenkomst met werkelijke verwijs- en voorschrijfgegevens;
 - betrouwbaarheid beoordelingsprocedures hoog (resp. $r = 0,83$ en $r = 0,89$).

Conclusie : Meetinstrumenten gezamenlijk discrimineren valide en betrouwbaar tussen huisartsen die 'veel' en huisartsen die 'weinig' kans op onnodige schade induceren.

Betrouwbaarh. : Zie resultaten.

Validiteit : - Subjectief hoog;
 - overeenkomst tussen gevonden factoren in dit onderzoek en dat van Mokkink.

- Auteur** : LoGerfo e.a., 1978.
- Doel** : Vergelijken van kwaliteit van zorg aan patiënten met urineweginfectie in een 'prepaid group practice' en in een 'independent practice setting'.
- Populatie** : 98 patiënten met urineweginfectie uit PGP en 69 patiënten met urineweginfectie uit IPP, geïdentificeerd door inspectie statussen over bepaalde periode (feb. '71 - juni '73) en met behulp van 'claims forms'.
- Aandoeningen** : Urineweginfectie.
- Handelingen** : Anamnese, lichamelijk onderzoek, laboratoriumonderzoek, therapie, follow-up.
- Bron van inf.** : Uittreksels statussen.
- Criteria** : Expliciete 'optimum' criteria, vastgesteld door panel van artsen uit de PGP en IPP.
- Resultaten** : Gemiddelde physician performance index (PPI) over beide groepen: 46,2 (s.d. = 19,4), max.: 100; PPI in PGP hoger ($p < 0,05$), met name wat betreft laboratoriumonderzoek.
- Conclusie** : De organisatie van een praktijk beïnvloedt het zorgproces in belangrijke mate.
- Betrouwbaarh.** : Geen test uitgevoerd.
- Validiteit** : Wijze van opstellen criteria (door 'peers') maakt een zeer valide indruk; echter alleen medisch-technische zorg geëvalueerd, geen verband gelegd met resultaat van zorg.

Auteur : Mansfield, 1986.

Doel : Nagaan hoe goed statussen worden bijgehouden door huisartsen.

Populatie : Overdrachtsgegevens naar een bepaalde praktijk van 100 verschillende huisartsen.

Aandoeningen : Niet van toepassing.

Handelingen : Registratie van persoonsgegevens en bepaalde medische gegevens.

Bron van inf. : Overdrachtsgegevens, statussen.

Criteria : - Expliciete beschrijving van resultaten, geen duidelijke vaststelling van wat wel/niet acceptabel is;
- impliciet oordeel over leesbaarheid.

Resultaten : - Sexe en leeftijd altijd geregistreerd, veel minder: burgerlijke staat, beroep, medische gegevens;
- leesbaarheid goed in 86,2%.

Conclusie : Verbetering is mogelijk.

Betrouwbaarh. : Geen betrouwbaarheidstest beschreven.

Validiteit : Geen commentaar auteurs.

Auteur : Michael, 1984.

Doel : Nagaan hoe goed de registratie is van bloeddruk in huisartspraktijken.

Populatie : 6 praktijken met opleiding in 'outer London', 6 gematchte praktijken in 'outer London' en 6 niet-opleidingspraktijken in 'inner London'; weigering in drie praktijken; 2.374 patiëntenstatussen.

Aandoeningen : Hypertensie.

Handelingen : Case-finding, registratie, behandeling.

Bron van inf. : Statussen.

Criteria : Expliciete beschrijving van bevindingen omtrent registratie en behandeling en vergelijking tussen praktijken zonder duidelijke vaststelling van wat wel en niet acceptabel is.

Resultaten : Beschrijvend; enige vergelijking: meer registratie in outer London ($p < 0,05$).

Conclusie : Registratie moet verbeterd worden.

Betrouwbaarh. : Geen commentaar auteur.

Validiteit : Geen commentaar auteur.

Auteur	: Makkink e.a., 1986.
Doel	: Nagaan of verschillende werkstijlen bij huisartsen te onderscheiden zijn en of deze samenhangen met bepaalde attitudes.
Populatie	: 57 huisartsen die meededen aan project 'Huisarts en somatische fixatie'.
Aandoeningen	: 24 protocolklachten.
Handelingen	: - Openstaan voor inbreng van patiënt; - expliciteren van het eigen handelen; - adequaat somatisch handelen; - adequaat psychosociaal handelen; - verwijsgedrag; - specifiek voorschrijfgedrag (pijnstillers, hoestdempers, psychofarmaca, antibiotica en antihistaminica).
Bron van inf.	: - Niet-participerende observatie; - audio-opnamen; - ziekenfondsgegevens; - vragenlijsten over attitudes.
Criteria	: 24 medisch-technische protocollen (expliciete criteria); beoordelingsinstrument voor voorwaardelijke en inhoudelijke vaardigheden.
Resultaten	: 43 van 55 huisartsen (bij twee huisartsen gegevens incompleet) waren in te delen in model van vier onderscheiden werkstijlen (geïntegreerd, doener, minimum + recept, minimum + verwijskaart). Verschillende werkstijlen corresponderen duidelijk met verschillen in opvatting over controle, risico's en medisch-technische taken.
Conclusie	: Er zijn verschillende werkstijlen te onderscheiden, maar aangezien hiervoor gewerkt wordt met relatieve criteria kan dit niet zonder meer op elke huisarts toegepast worden; het model dient nog nader getoetst te worden; samenhang met attitudes bevestigt de realiteitswaarde van het model; het model betekent een nuancering van de Nijmeegse protocollen en normen.
Betrouwbaarh.	: Elders beschreven.
Validiteit	: Zie conclusies, ook elders beschreven.

- Auteur** : Norman e.a., 1985.
- Doel** : Onderzocht worden methoden om criteria vast te stellen en verschillen tussen huisartsen met betrekking tot procedures en verslaglegging worden gemeten met behulp van simulatiepatiënten.
- Populatie** : 8 'family physicians' die deelnamen aan teams voor criteriasetting en 7 andere 'family physicians' (vrijwilligers).
- Aandoeningen** : 7 uitvoerig vastgelegde casussen.
- Handelingen** : Alle verrichte handelingen (zowel medisch-technisch als psychosociaal).
- Bron van inf.** : - Recall-formulier ingevuld door simulatiepatiënt;
- Status.
- Criteria** : Vastgesteld door team van 'family physicians' en specialist (expliciet).
Vier maten: totaalscore (alle criteria door één van de teamleden voorgesteld); gewogen score (gewicht gebaseerd op aantal teamleden dat criteria voorstelt); specialistenscore; consensusscore (alle criteria door alle 4 family physicians genoemd).
- Resultaten** : - Correlaties tussen de 4 scoringsmethoden redelijk hoog (totaal-consensus 0,69, totaal gewogen 0,96);
- geen significante verschillen tussen groep die wel en groep die geen criteria vastgesteld heeft;
- door artsen die betrokken waren bij criteria setting werd op dezelfde casussen als volgt gescoord: circa 0,5 van de totaalscore en circa 0,75 van de specialisten en consensusscore; verslaglegging bedroeg respectievelijk 0,33 en 0,5.
- Conclusie** : De methode met simulatiepatiënten is haalbaar en bruikbaar; criteria kunnen beter gebaseerd worden op het handelen in werkelijkheid dan op een geïdealiseerd beeld.
- Betrouwbaarh.** : Interraterbetrouwbaarheid recall formulieren hoog: $K = 0,86$.
- Validiteit** : Simulatiepatiënten meestal niet ontdekt; criteria lijken niet erg valide; niet vinden van verschil tussen bij criteria setting betrokken artsen en overigen mogelijk door te weinig casussen of door selecte onderzoeksgroep; voor generaliseerbare resultaten circa 30 simulatiecasussen per arts nodig.

Auteur	: Novick e.a., 1976.
Doel	: Vergelijken van de resultaten van meting van de kwaliteit van zorg in een pediatrische polikliniek met behulp van de tracertechniek met die van een impliciete kwaliteitsbeoordeling door specialisten.
Populatie	: Statussen van 100 patiëntjes, minimaal 6 maanden oud, waarbij Hb-bepaling was aangevraagd en de uitslag lager of gelijk was aan 11g/100ml (na uitsluiting van chronische aandoeningen).
Aandoeningen	: IJzergebreksanemie.
Handelingen	: Anamnese en onderzoek, diagnostiek, behandeling, follow-up.
Bron van inf.	: Statussen.
Criteria	: 1. Expliciete criteria, ontwikkeld door een panel van specialisten, daarna voorgelegd aan alle beoordeelde artsen, 80% consensus vereist, relatief gewicht bepaald door panel van 3 specialisten, toegepast door niet-medici. 2. Impliciet oordeel van 4 specialisten (geen aanwijzingen), wel/niet acceptabel.
Resultaten	: 1. In slechts 6% van de 48 gevallen waarbij anemie werd ontdekt, bedroeg de kwaliteitsscore 81-100%. 2. 33,3% van de 48 herkende gevallen: niet acceptabel voor 2 specialisten; 27,1% acceptabel voor één van beide, 37,5% acceptabel voor beiden. Correlatiecoëfficiënt tussen expliciet en impliciet oordeel 0,37-0,46 ($p < 0,05$).
Conclusie	: De methode is haalbaar en identificeert tekortkomingen die te verhelpen zijn.
Betrouwbaarh.	: - Geen commentaar over betrouwbaarheid niet-medische beoordelaars; - circa 70% overeenstemming tussen impliciete beoordelaars.
Validiteit	: Methode van criteria setting maakt valide indruk, nog versterkt door correlatie met impliciet oordeel.

- Auteur** : Palmer e.a., 1984.
- Doel** : Evaluatie van de kwaliteit van de hulpverlening in ambulante medische zorg.
- Populatie** : Twee opleidingsziekenhuizen en 6 gezondheidscentra.
- Aandoeningen** : Lage hematocriet, cervixca, mamma, indicaties voor digoxine, verhoogd bloedglucosegehalte.
- Handelingen** : Preventie, aanvullende diagnostiek, registratie, therapie, follow-up.
- Bron van inf.** : Statussen.
- Criteria** : Expliciet, vertakt, ongewogen, opgesteld door hulpverleners zelf; afwijkingen beoordeeld door een afgevaardigde van de hulpverleners als al dan niet gerechtvaardigd.
- Resultaten** : Percentage casussen met één of meer ongerechtvaardigde afwijkingen: bij lage hematocriet 31%, bij screening 31%, bij verhoogd glucose 6% en bij digoxine 42% (gemiddeld aantal criteria respectievelijk 2, 2, 1,5 en 9,5).
- Conclusie** : Methode is bruikbaar, nogal wat variatie tussen instellingen.
- Betrouwbaarh.** : Geen interraterbetrouwbaarheid bepaald; waarschijnlijk niet blind beoordeeld; er lijkt nogal wat verschil te bestaan tussen arts-beoordelaars. 3-5% van de afwijkingen gevonden door niet-medici berustten op fouten in dataverzameling.
- Validiteit** : Criteria worden geacht van belang te zijn voor resultaat bij de patiënt; niet duidelijk in hoeverre dit zo is; probleem van verschil in belangrijkheid/geen verschillend gewicht.

- Auteur** : Payne e.a., 1984.
- Doel** : Verband nagaan tussen eigenschappen van artsen (waarbij duur van de opleiding, leeftijd, certificatie en dergelijke) en kwaliteit van hun handelen bij ambulante zorg.
- Populatie** : 5 soorten praktijken: solopraktijken, universiteitscentrum, geaffilieerde polikliniek, prepaid polikliniek; per praktijksoort 150 casussen per diagnose.
- Aandoeningen** : Indicaties voor chlooramphenicol, keflex, digitalis en prednison, anemie, essentiële hypertensie, chronische hartaandoeningen, vulvovaginitis, acute urineweginfectie, chronische/intermitterende urineweginfectie.
- Handelingen** : Preventieve, acute en chronische zorgaspecten.
- Bron van inf.** : Uittreksels van statussen.
- Criteria** : Expliciete optimum-criteria voor process en outcome, ontwikkeld door artsen uit alle verschillende soorten praktijken, gewogen op basis van consensus.
- Resultaten** : Geen consistent verband tussen certificatie en kwaliteit.
Inconsistent patroon van iets betere verrichtingen bij artsen die korter geleden afgestudeerd zijn.
Redelijk consistent beeld van betere kwaliteit van specialisten op eigen vakgebied.
- Conclusie** : Kwaliteit hangt in volgorde van belangrijkheid samen met: vakgebied, aantal jaren sinds afstuderen, certificatie.
- Betrouwbaarh.** : Geen test beschreven.
- Validiteit** : Elders beschreven.

Auteur : **Plaisier, 1987.**

Doel : Vergelijken van handelwijze van huisartsen.

Populatie : Vijf huisartsen.

Aandoeningen : Niet van toepassing.

Handelingen : Handelen tijdens spreekuur.

Bron van inf. : Directe, niet-participerende observatie.

Criteria : Impliciet; opvallende zaken worden genoteerd.

Resultaten : Beschrijving van opvallende zaken, gerelateerd aan tijdsduur van verschillende fasen van de consulten: veel variatie.

Conclusie : Op deze wijze kan men in korte tijd veel informatie verkrijgen over de werkwijze van een (collega)huisarts.

Betrouwbaarh. : Eén observator.

Validiteit : Geen commentaar auteur.

- Auteur** : Van Ree e.a., 1985.
- Doel** : Het verband tussen het resultaat van hypertensie-behandeling op langere termijn en de praktijkvoering wordt nagegaan.
- Populatie** : Vier huisartsenpraktijken;
186 patiënten uit de interventiegroep, 92 patiënten uit de controlegroep.
- Aandoeningen** : Hypertensie (mannen jonger dan 40 jaar: 90 mmHg, ouder dan 40 jaar: 95 mmHg; vrouwen jonger dan 40 jaar 95 mmHg, ouder dan 40 jaar 100 mmHg).
- Handelingen** : Werkwijze ten aanzien van hypertensie-patiënten.
- Bron van inf.** : - Meting bij patiënten;
- interview met huisartsen en assistentes over werkwijze;
- interview met patiënten over adviezen en controlefrequentie;
- statussen.
- Criteria** : Expliciete beschrijving van resultaten; geen duidelijke vaststelling van wat wel en niet acceptabel is.
- Resultaten** : Aantal patiënten dat streefwaarde heeft gehaald of gehandhaafd duidelijk verschillend tussen praktijken; verschil hangt samen met de werkwijze in deze praktijken (namelijk bewaking van controlefrequentie); ook samenhang met door patiënt vermelde regelmatige controles en met geregistreerde controles op statussen.
- Conclusie** : De huisarts kan een belangrijke bijdrage leveren aan de behandeling van hypertensie door te zorgen voor een systematische praktijkvoering.
- Betrouwbaarh.** : Geen test beschreven; mogelijk elders beschreven.
- Validiteit** : Verschillende informatiebronnen wijzen in dezelfde richting.

Auteur	: Rethans en Van Boven, 1987.
Doel	: Vergelijken van indirecte beoordeling van medisch handelen (vragenlijst) met directe methode (simulatiepatiënten).
Populatie	: 25 huisartsen.
Aandoeningen/ Handelingen	: Medisch handelen bij urineweginfectie.
Bron van inf.	: 1. Beoordeling door simulatiepatiënt; 2. beantwoording vragenlijst.
Criteria	: Protocol Nijmeegs Universitair Huisartsen Instituut (consensus, expliciete criteria).
Resultaten	: Wat betreft obligate handelingen doen huisartsen bij de simulatie meer (64%) dan zij aangeven bij de vragenlijst; voor facultatieve en overbodige handelingen is dit omgekeerd.
Conclusie	: - De methode van simulatiepatiënten geeft meer relevante informatie dan een indirecte methode; - het zou beter zijn empirische standaarden vast te stellen dan consensusstandaarden (deze laatste komen niet genoeg met de werkelijkheid overeen).
Betrouwbaarh.	: Tussen beoordelaars, evenals test-hertest betrouwbaarheid hoog ($\geq 0,85$).
Validiteit	: Er wordt aannemelijk gemaakt dat de methode met simulatiepatiënten meer valide is dan de indirecte methode.

Auteur : Romm en Putnam, 1981.

Doel : Bepalen van de overeenkomst tussen verslaglegging in de status en inhoud van het spreekuurcontact.

Populatie : 22 artsen in 'primary care', 55 spreekuurcontacten.

Aandoeningen : Alle voorkomende.

Handelingen : Verslaglegging en verbale communicatie.

Bron van inf. : 1. Woordelijke weergave spreekuurcontacten;
2. statussen.

Criteria : Geen criteria vastgelegd; overeenkomst bepaald.

Resultaten : 59% van de items uit één van beide bronnen werd teruggevonden in de status;
voor de hoofdklacht: 92%;
huidige klachten : 71%;
rest anamnese : 29%.

Conclusie : De status voldoet niet als informatiebron omtrent de inhoud van het arts-patiënt-contact.

Betrouwbaarh. : Interbeoordelaarbetrouwbaarheid wordt hoog genoemd; echter de opzet was niet volledig blind.

Validiteit : De validiteit van de weergave is hoog, die van de status wordt onderzocht en blijkt niet voldoende.

- Auteur** : Rudnick, 1983.
- Doel** : Nagaan of een educatief programma voor huisartsen effect heeft op de kwaliteit van de geleverde zorg.
- Populatie** : 8 huisartsen in een experimentele groep, 8 in een controlegroep; meer dan 4.500 patiëntencontacten.
- Aandoeningen** : 6 van 18 veelvoorkomende aandoeningen; 2 naar keuze en 2 verplicht en 2 'verborgen' indicator aandoeningen.
- Handelingen** : Handelwijze tijdens spreekuur.
- Bron van inf.** : Statussen; uittreksels gemaakt door nurse-abstractors (blind).
- Criteria** : Expliciet, ontworpen door peer advisory groep; beschreven als 'superior', 'adequate' of 'intermediate'.
- Resultaten** : Beide groepen waren gelijk wat betreft kwaliteit vóór toepassing van het programma, bij alle 6 indicatoraandoeningen; ook na het programma geen significant verschil tussen beide groepen.
- aandoeningen naar keuze: beide groepen 6% verhoging 'adequate' of 'superior';
 - verplichte aandoeningen: experimentele groep 10% verhoging, controlegroep 1% verlaging (verschil significant);
 - 'verborgen' aandoeningen: in beide groepen niet-significante verlaging.
- Conclusie** : Kwaliteit verbetert alleen bij nascholing op verplichte aandoeningen; géén uitstralingseffect van nascholing.
- Betrouwbaarh.** : 'Nurse-abstractors', 'blind', intra- en interobserver betrouwbaarheid, $K > 0,8$.
- Validiteit** : Procesbeoordeling van indicatoraandoeningen staat wel bloot aan kritiek. Gevoeligheid van de methode lijkt groot op basis van kleine gevonden significante verschillen bij voor- en nameting.

Auteur : Sheps en Robertson, 1984.

Doel : Evaluatie van de zorg van een bepaalde multidisciplinary community family practice clinic.

Populatie : 5 part-time artsen, 124 zorgepisoden bij 103 patiënten.

Aandoeningen/ Handelingen : - Otitis media;
- hypertensie;
- prenatale zorg;
- zuigelingenverzorging;
- inentingen tot leeftijd van 24 maanden;
- depressie;
- urineweginfectie.

Bron van inf. : Patiëntenkaarten werden beoordeeld met behulp van expliciete criteria uit Burlington Randomised Controlled Trial door niet-medicus.

Criteria : Elders beschreven.

Resultaten : 67% van alle zorgepisoden is adequaat of beter dan dat (variërend van 33% voor hypertensie tot 81% voor zuigelingenverzorging).

Conclusie : Onderzoeksmethode is bruikbaar voor evaluatie van zorg, bijvoorbeeld om effect van nascholing te meten.

Betrouwbaarh. : Overeenstemming tussen beoordelaars 85%.

Validiteit : Verwezen wordt naar de Burlington Randomised Controlled Trial.

Auteur : Sibley e.a., 1975.

Doel : Toepasbaarheid, betrouwbaarheid en validiteit beoordelen van een nieuw in te voeren meetinstrument voor de kwaliteit van zorg.

Populatie : Drie verschillende praktijken.

Aandoeningen : 10 indicator-aandoeningen.

Handelingen : Voorschrijfbeleid bij 13 geneesmiddelen; bij 10 indicator-aandoeningen verwijsbeleid.

Bron van inf. : - Patiëntenkaarten;
- vragenlijsten aan specialisten naar wie verwezen wordt;

Criteria : Expliciete criteria, opgesteld door 'Peer Advisory Group', onbekend aan beoordeelde artsen.

Resultaten : Variatie tussen praktijken en tussen aandoeningen/geneesmiddelen.

Conclusie : Goede methode die verder toegepast kan worden.

Betrouwbaarh. : Interrater overeenstemming minimaal 88%.

Validiteit : Door 'Peer Advisory Group' in wier praktijken methode werd uitgetest, hoog geacht.

- Auteur** : Simpson e.a., 1987.
- Doel** : Vergelijken van het diagnostisch proces van 'family physicians' en 'general internists'.
- Populatie** : 12 'internists' en 12 'family physicians' in primary care; in elke groep 4 1^e jaars-assistenten, 4 3^e jaars-assistenten en 4 ervaren artsen.
- Aandoeningen** : Drie simulatiecasussen op papier:
- dyspnoe + pijn op de borst;
- buikpijn;
- syncope.
- Handelingen** : Hypothese-vorming bij diagnostiek.
- Bron van inf.** : Audio-opnamen van 'hardop denken' tijdens lezen van kaarten over casussen.
- Criteria** : Niet vastgelegd wat al dan niet acceptabel is; zuiver vergelijkend onderzoek; wel van alle casussen uiteindelijk diagnose vastgelegd.
- Resultaten** : - Family physicians uiten minder hypothesen dan internisten;
- hypothesen van internisten zijn meer specifiek; geen verschil wat betreft correctheid uiteindelijke diagnose.
- Conclusie** : Internisten en family physicians hebben een verschillende werkwijze.
- Betrouwbaarh.** : Beoordelaars getraind tot minstens 80% overeenstemming.
Generaliseerbaarheid beperkt (3 casussen, alle artsen van dezelfde universiteit).
- Validiteit** : 'This study cannot answer the important question of who is the better primary care physician'.

Auteur : Singh, 1984.

Doel : Vergelijken van de mate van metabole controle bij diabetes-patiënten uit een huisartsenpraktijk met die bij patiënten uit een ziekenhuispolikliniek.

Populatie : Aselecte steekproef uit de huisartsenpraktijk: 51 patiënten met alleen dieet, 64 met orale antidiabetica, 48 met 1 maal daags insuline, 58 met tweemaal daags insuline; gematchte groep uit ziekenhuis (leeftijd, sexe, duur van de diabetes en gewicht).

Aandoeningen : Diabetes.

Handelingen : Metabole controle, aanvragen en registreren van laboratoriumtesten.

Criteria : Niet vastgelegd wat wel of niet acceptabel is; vergelijkingen gemaakt tussen twee groepen.

Resultaten : Geen significant verschil in bloedsuiker of HbA1.

Conclusie : Huisartsen kunnen even goed metabole controle bereiken bij diabetes mellitus als ziekenhuizen.

Betrouwbaarh. : Geen test beschreven.

Validiteit : Bloedsuiker is slechts één van de indicatoren van goede zorg voor diabetes patiënten.

Auteur : Smith, 1977.

Doel : Beoordeling van de kwaliteit van zorg in een bepaalde 'family practice' aan de hand van een tracer-aandoening: hypertensie.

Populatie : Aselecte steekproef van 10% van de statussen van patiënten met hypertensie in één 'family practice' (45 statussen).

Aandoeningen : Hypertensie.

Handelingen : Anamnese, onderzoek, laboratorium, diagnostiek, behandeling.

Criteria : Expliciete minimumcriteria vastgesteld door het 'Institute of Medicine'.

Resultaten : In een aselecte steekproef van 100 statussen, één patiënt met geregistreerde hoge bloeddruk niet gediagnosticeerd; twee van de 45 geselecteerde hypertensie-patiënten fout-positief; voldoen aan de criteria varieert van 78% (behandeling) tot 95% (lichamelijk onderzoek).

Conclusie : De methode is veelbelovend en haalbaar in 'family practice'.

Betrouwbaarh. : Geen commentaar van de auteur.

Validiteit : Onduidelijk.

Auteur : Smits e.a., 1985.

Doel : Verband nagaan tussen patiëntgerichte werkwijze en adequaat diagnostisch en therapeutisch beleid.

Populatie : 815 spreekuurcontacten van 57 huisartsen.

Aandoeningen : 24 protocolklachten.

Handelingen : Gedragswetenschappelijke en medisch-technische vaardigheden.

Bron van inf. : Audio-opnamen.

Criteria : - Protocollen voor medisch-technisch handelen; expliciet, voor obligate en overbodige handelingen.
- Beoordelingsschalen voor gedragswetenschappelijke vaardigheden.

Resultaten : - Correlatiecoëfficiënten tussen 'openstaan voor de inbreng van de patiënt' en 'expliciteren van het eigen handelen' enerzijds en 'psychosociale aandacht', 'adequate diagnostiek' en 'adequate therapie' anderzijds variëren tussen 0,41 en 0,57.

Conclusie : Huisartsen die zich meer openstellen voor de inbreng van de patiënt en het eigen handelen meer expliciteren, reageren adequater op klachten, zowel in medisch-technisch als in psychosociaal opzicht.

Betrouwbaarh. : - Medische en gedragswetenschappelijke aspecten door verschillende beoordelaars gescoord;
- medisch-technisch deel: interreliability-coëfficiënt 0,70-0,90;
- gedragswetenschappelijk deel: interreliability-coëfficiënt 0,30-0,60.

Validiteit : Protocollen uitgebreid onderbouwd.

Auteur : Taylor, 1987.

Doel : Nagaan of een programma om de behandeling van epilepsie te verbeteren in een 'general practice' effect heeft gehad.

Populatie : Alle epilepsie-patiënten (ook onder specialistische behandeling) in een bepaalde praktijk met vier general practitioners (n = 34 in 1980; n = 46 in 1986).

Aandoeningen : Epilepsie (duidelijk gedefinieerd).

Handelingen : Diagnostiek, therapie en begeleiding.

Bron van inf. : Statussen (?).

Criteria : Niet duidelijk beschreven; beoordeeld worden: aantal toevallen, medicamenteuze therapie, veranderingen in therapie.

Resultaten : Beschrijvend.

Conclusie : Het programma blijkt werkzaam; de essentie hiervan is nauw overleg tussen arts, patiënt en familie.

Betrouwbaarh. : Geen commentaar.

Validiteit : Geen commentaar.

- Auteur** : Thomassen, 1986.
- Doel** : Verband nagaan tussen communicatieve en kwalitatieve aspecten van het huisarts-patiënt-contact.
- Populatie** : 315 spreekuurcontacten (nieuwe klachten) in 52 huisartspraktijken (niet aselekt).
- Aandoeningen** : Alle voorkomende klachten.
- Handelingen** : Communicatieve vaardigheden.
- Bron van inf.** : Letterlijke vastlegging van verbale informatie uit consult.
- Criteria** : Semi-expliciet; beoordeeld door onderzoekers.
- Resultaten** : Communautiteit (mate van gemeenschappelijkheid in de preoccupaties van arts en patiënt met zaken die voor hen van belang zijn in het gesprek) en kwaliteit hangen wel met elkaar samen; geen sterk of eenduidig verband.
- Conclusie** : 'Hoe meer communautiteit in het gesprek is toe te schrijven aan geslaagde informatie-uitwisseling op geleide van zowel de huisarts als de patiënt, hoe meer de geboden hulp is afgestemd op geneeskundige en persoonsgebonden facetten van de problematiek van de patiënt.'
- Betrouwbaarh.** : (Meetinstrument voor kwaliteit) consulten steeds beoordeeld door 3 onderzoekers van verschillende disciplines; in 71% volledige overeenstemming, in 21% overeenstemming tussen 2 van de 3, en in 8% geen overeenstemming.
- Validiteit** : Bij 65 consulten tevens impliciet oordeel van huisarts-deskundige over kwaliteit (interbeoordelaarbetrouwbaarheid hoog: 93% overeenstemming); goede samenhang met kwaliteit zoals eerder bepaald is ($r = 0,37$; $p = 0,002$).

- Auteur** : **Thompson en Osborne, 1976.**
- Doel** : Nagaan of statussen voldoende informatie bevatten om de kwaliteit van zorg te beoordelen aan de hand van 144 criteria.
- Populatie** : 10.500 statussen van 166 artsen (kinderartsen en family physicians) (niet representatief).
- Aandoeningen** : Gezonde kinderen, tonsillo-pharyngitis, asthma bronchiale, urineweginfectie bij meisjes.
- Handelingen** : Sympervisie gezonde kinderen, anamnese, lichamelijk onderzoek, laboratoriumonderzoek, begeleiding bij drie aandoeningen.
- Bron van inf.** : Statussen.
- Criteria** : 144 expliciete criteria, vastgesteld door een commissie, gevalideerd door experts en door een aselechte steekproef van artsen uit heel de V.S.
- Resultaten** : - Slechts 24 van de 144 criteria waren geregistreerd in 75% of meer van alle statussen;
- slechts 48% van de artsen vond dat de methode een juist beeld gaf van hun handelen.
- Conclusie** : De methode is op zich bruikbaar, maar de registratie van artsen maakt de methode onbruikbaar voor het beoordelen van de kwaliteit van het handelen.
- Betrouwbaarh.** : De arts beoordeelde de bevindingen van de reviewer; voor 77 van de 144 criteria was in alle gevallen overeenstemming; voor de overige 67 criteria was in minstens 96% van de gevallen overeenstemming.
- Validiteit** : Ondanks uitgebreide validering vooraf, vond slechts 48% van de artsen dat hun werk goed werd weergegeven door de methode.

- Auteur** : Tielens e.a., 1986.
- Doel** : Nagaan of er verschillen in kwaliteit bestaan bij het medisch handelen bij vrouwelijke en mannelijke patiënten.
- Populatie** : 56 mannelijke huisartsen; 955 klachten, verdeeld over 16 protocolklachten (niet geslachtsspecifiek); 529 (55%) bij vrouwen; 426 (45%) bij mannen.
- Aandoeningen** : 16 protocolklachten.
- Handelingen** : Medisch handelen.
- Bron van inf.** : Audio-opnamen en niet-participerende observatie.
- Criteria** : Expliciete protocollen obligate/overbodige handelingen opgesteld door 20 huisartsen.
- Verschil in kwaliteit bij mannelijke en vrouwelijke patiënten vrijwel nihil ($\delta < 0,9$);
- tussen individuele contacten grote verschillen die met andere factoren dan het geslacht van de patiënt samenhangen;
- bij verschillende klachten/verschillende onderdelen van het consult wel geringe verschillen.
- Conclusie** : Verschillen in kwaliteit bij medisch-technisch handelen van mannelijke huisartsen bij mannelijke en vrouwelijke patiënten zijn klein en niet van praktische klinische betekenis.
- Betrouwbaarh.** : Elders beschreven; hoog.
- Validiteit** : Idem.

Auteur : Wilson, 1987.

Doel : Onderzoeken in hoeverre huisartsen het screeningsprogramma op cervixcarcinoom uitvoeren voor vrouwen tussen 35 en 64 jaar.

Populatie : 240 huisartsen, verbonden aan de Vale of Trent faculty; response 101 (42%); 7.600 statussen.

Aandoeningen : Cervixcarcinoom.

Handelingen : Screening.

Bron van inf. : Vragenlijst over beleid; statussen.

Criteria : Geen vaststelling van wat wel/niet acceptabel is.

Resultaten : - Grote variatie in gemeld beleid: 47% zegt een normaal uitstrijkje na 1 jaar te willen herhalen; 18% na 5 jaar;
- 46% van de praktijken heeft een actief oproepsysteem;
- percentage vrouwen 35-64 die in de laatste 5 jaar een uitstrijkje heeft gehad: 20-86%;
- percentage vrouwen 35-46 van wie geen enkel uitstrijkje geregistreerd staat: 6-51%.

Conclusie : Ook goed gemotiveerde praktijken met een duidelijk beleid kunnen onvoldoende effectief screenen.

Betrouwbaarh. : Twijfelachtig; door artsen zelf uitgevoerd, statussen vormen geen aselechte steekproef.

Validiteit : Geen commentaar auteurs.

- Auteur** : **Wilson & Murray, 1985.**
- Doel** : Nagaan hoe de kwaliteit is van de behandeling van Parkinson patiënten door huisartsen.
- Populatie** : Alle Parkinson patiënten die gevonden werden in een bepaald gebied met 26 huisartsen en circa 57.000 patiënten: 32 patiënten.
- Aandoeningen** : Ziekte van Parkinson.
- Handelingen** : Medicamenteuze therapie, verwijzing, diagnostiek.
- Bron van inf.** : Statussen.
- Criteria** : Niet vastgesteld wat wel en niet acceptabel is.
- Resultaten** : Beschrijvend; slechts vier van 19 patiënten behandeld met Sinemet, kregen volgens farmacologische begrippen voldoende carbidopa;
Alle zeven patiënten die nooit verwezen waren, hadden minimaal 2 'major' en 1 'minor' of 1 'major' en 2 'minor' symptomen.
- Conclusie** : Weinig van de meest recente ontwikkelingen worden weerspiegeld in de behandeling in de huisartsenpraktijk.
- Betrouwbaarh.** : Geen commentaar auteurs.
- Validiteit** : Nergens outcome-maten gehanteerd (HR).

- Auteur** : Van Weel en Van Zelst, 1983.
- Doel** : Effect nagaan van op het handelen van de huisarts (wat betreft diabetes) gericht onderzoek dat eerder werd uitgevoerd.
- Populatie** : Alle ingeschreven patiënten van één praktijk in een gezondheidscentrum.
- Aandoeningen** : Diabetes mellitus.
- Handelingen** : Registratie, verwijzing, therapie, controle instelling, bewaken chronische complicaties.
- Bron van inf.** : Patiëntenkaarten, archief.
- Criteria** : 'Schema diabetes mellitus' van het NHI; weergave van resultaten en vergelijking met eerder onderzoek; niet gespecificeerd wat wel en wat niet acceptabel is.
- Resultaten** : - Van 81 patiënten met diagnose DM op de kaart 8 onjuist;
- van 12 insuline-afhankelijke patiënten 9 bij internist; van 56 niet-insuline afhankelijke patiënten 2 bij internist;
- scherpte van instelling geringer dan bij eerste onderzoek;
- bij 79% van de patiënten is ooit funduscopie verricht; optreden van complicaties wordt bij een groter deel van de patiënten bewaakt dan eerst.
- Conclusie** : Meer patiënten door de huisarts behandeld, minder scherp ingesteld, meer screening, maar voor het grootste deel zonder therapeutische consequenties; vraagtekens worden geplaatst bij de bruikbaarheid van het schema.
- Betrouwbaarh.** : Geen commentaar auteurs.
- Validiteit** : Validiteit van het gehanteerde schema in twijfel getrokken; wat is het belang voor de patiënt bij allerlei voorschriften en protocollen?

- Auteur** : Zuckerman e.a., 1980.
- Doel** : Meten van kwaliteit van zorg in een primary care clinic met behulp van de trajectory methode.
- Populatie** : 72 patiënten die een afspraak hadden voor Rö-foto's op gastrointestinaal gebied in 1 kliniek.
- Aandoeningen/ Handelingen** : Beleid bij acute gastrointestinale klachten.
- Bron van inf.** : - Patiëntenstatus;
- Telefonisch interview met patiënt;
- Diagnose radioloog.
- Criteria** : Expliciete criteria voor medisch-technische zorg en voor arts-patiëntrelatie.
- Resultaten** : 50% van de patiënten ontving effectieve zorg.
- Conclusie** : Ondanks problemen in de methode (bijvoorbeeld het niet betrekken van uitwendige omstandigheden die van invloed zijn op outcome) heeft deze ook goede kanten, zoals het verband tussen process en outcome dat onderzocht kan worden.
- Betrouwbaarh.** : Hiervan wordt geen test vermeld.
- Validiteit** : De criteria die waren overgenomen van een ander onderzoek werden door 2 artsen van het ziekenhuis valide geacht. Er bleek echter toch geen consensus te zijn over de criteria in de onderzochte groep.

