

Registratie in de huisartspraktijk

Over de betrouwbaarheid, praktische hanteerbaarheid en adequaatheid van het RFE-classificatiesysteem

F. VAN DER HORST, A. SELEN, F. VISSERS, H. PLAGGE,
R. VON HÖFEN, DR. C. DE GEUS EN J. KLEIJNEN*

Acht huisartsen hebben gedurende drie aaneengesloten maanden alle redenen voor contact met patiënten vastgelegd en geclassificeerd. De classificatie vond plaats volgens een proefversie van de Reason for Encounter-lijst (RFE-C). In dit artikel wordt de betrouwbaarheid van de classificatie besproken. Daarnaast wordt aandacht geschonken aan adequaatheid en praktische hanteerbaarheid van het classificatie-instrument. Geconcludeerd wordt dat de betrouwbaarheid goed is. Proefondervindelijk is gebleken dat het systeem voor de deelnemende huisartsen adequaat en praktisch hanteerbaar is. Voor full-time in de praktijk werkende huisartsen blijkt het zeer wel mogelijk wetenschappelijk verantwoord te registreren met behulp van de RFE-classificatie.

Inleiding

Registratie aan de hand van classificatiesystemen in de huisartspraktijk is niet nieuw. Bekende systemen voor de huisarts zijn de uit 1963 daterende E-lijst (Eimerl) en de uit 1979 daterende tweede versie van de International Classification of Health Problems in Primary Care (ICHPPC-2); beide zijn afgeleid van de International Classification of Diseases (ICD). De ICHPPC-lijst onderscheidt zich van de E-lijst door een sterke probleemoriëntatie. Dit laatste geldt nog meer voor de Reason for Encounter Classification (RFE-C).

Probleemgeoriënteerd registreren komt vooral dan tot zijn recht als de registratie aansluit op *probleemgeoriënteerde verslaglegging* (Problem Oriented medical Record of POR¹).

In dit verslagleggingssysteem staat de SOEP-notatie van het journaal centraal. Deze notatie blijkt een uniek hulpmiddel te zijn om het probleem-oplossend proces objectieverbaar te maken. Op de

SOEP-elementen geënte classificatiesystemen kunnen behulpzaam zijn bij het registreren van de onderscheiden elementen uit het probleem-oplossend proces van de huisarts.

Het systematisch en in hun onderlinge samenhang beschrijven van invoer (S), doorvoer (O en E) en uitvoer (P) van aan de huisarts aangeboden problemen vormt de doelstelling van het Basisproject Verslaglegging en Registratie dat aan de Rijksuniversiteit Limburg wordt uitgevoerd.² In dit artikel rapporteren we over onze ervaringen met de eerste stap in dit project, het registreren van de S (= subjectief). Acht huisartsen registreerden hiertoe gedurende de eerste drie maanden van 1983 alle redenen voor contact op basis van de door patiënten gegeven klachten, problemen of vragen. Dat gebeurde in het kader van een internationale *trial* van de WHO-werkgroep voor de International Classification for Primary Care (ICPC).

Over de opzet en inhoudelijke resultaten van deze *trial* zijn inmiddels verschillende publikaties verschenen.³⁻⁷ In de volgende bijdrage wordt ingegaan op de resultaten in termen van betrouwbaarheid (*reliability*), adequaatheid (*adequacy*) en praktische hanteerbaarheid (*feasibility*) (kader op bladzijde 230). De vraagstelling van het onderzoek luidt: in hoeverre is de RFE-C lijst een betrouwbaar instrument en vanuit het

gezichtspunt van de huisarts praktisch hanteerbaar en adequaat om de invoer (S) in termen van redenen voor contact met patiënten te registreren.

Methoden

Reason for Encounter-Classification

Over de Reason for Encounter Classification is in *Huisarts en Wetenschap* al eerder uitvoerig bericht;⁷ daarom wordt hier slechts kort op enkele hoofdzaken ingegaan.

Het is van belang zich te realiseren dat de RFE-C betrekking heeft op de redenen voor contact *zoals de huisarts deze percipieert*; de werkelijke contactredenen kan dus een geheel andere zijn.

De RFE-lijst is gebaseerd op twee assen. Op de horizontale as zijn zestien hoofdstukken uitgezet; dertien ervan hebben betrekking op orgaansystemen, terwijl drie hoofdstukken de psychische, sociale en algemene, niet gespecificeerde problematiek bevatten. Op de verticale as zijn zeven componenten uitgezet: Symptoms and complaints; Diagnostic, screening and preventive procedures; Treatment procedures and medications; Test results; Administrative; Other reasons; Diagnosis and diseases.

De tweede tot en met de zesde component zijn in alle hoofdstukken uit dezelfde rubrieken opgebouwd, terwijl de eerste en zevende component specifiek per hoofdstuk zijn. De componenten twee tot en met zes bevatten tezamen voor ieder hoofdstuk 33 codes. De eerste en zevende component zijn tamelijk uitgebreid, met in totaal respectievelijk 216 en 284 codes.

Elke code is opgebouwd uit drie delen (digits): de eerste verwijst met een letter naar het hoofdstuk, terwijl de laatste twee delen, bestaande uit twee cijfers, verwijzen naar de rubrieken binnen de componenten.

Overigens dient er op te worden gewezen dat de in dit onderzoek geteste classificatielijst niet meer wordt gebruikt. Deze lijst is – bijgesteld op grond van onze ervaringen en die van huisartsen-onderzoekers in andere landen – opgegaan in een nieuw systeem: de International Classification of Primary Care (ICPC – zie ook het bericht op pagina 246 van deze aflevering van *Huisarts en Wetenschap*).

Populatie

Aan de studie namen acht huisartsen deel, verdeeld over één solopraktijk (één huisarts), één associatiepraktijk (twee huisartsen) en drie gezondheidscentra (zes huisartsen). Hierbij dient

* F. van der Horst – socioloog, A. Seelen – huisarts, F. Vissers – huisarts, H. Plagge – student-assistent, R. von Höfen – systeemanalist, Dr. C. de Geus – huisarts, J. Kleijnen – onderzoeksassistent. Alle auteurs zijn betrokken bij het basisproject Verslaglegging en Registratie in de huisartspraktijk van de Rijksuniversiteit Limburg. Behalve de auteurs werken hier aan mee: E. Ebbens, Th. Govaert, L. Huynen, A. Knottnerus, R. Leclercq, J. Metsmakers, P. Muylers, M. van Nunen, H. J. Rop, allen huisarts. Projectleider: Prof. dr. C. A. de Geus.

opgemerkt te worden dat er geen gezondheidscentra bij waren waarvan alle huisartsen participeerden. De totale praktijkpopulatie van de deelnemende huisartsen bedroeg circa 22.000 patiënten. Onze veldoefening van drie maanden heeft betrekking op 10.045 patiëntencontacten en 15.195 redenen voor contact.

Door de verschillende praktijkvormen en het grote aantal redenen voor contact was de variatie in aanbod van klachten, problemen en vragen groot genoeg om de RFE-C op betrouwbaarheid, adequaatheid en praktische hanteerbaarheid te kunnen toetsen.

De voorbereiding

De voorbereidende activiteiten hadden als functie:

- duidelijkheid verkrijgen omtrent de systematiek van de classificatie en de (betekenis van de) instructie(lijst);
- globale toetsing van de adequaatheid van de RFE-lijst;
- globale toetsing van de praktische hanteerbaarheid van de RFE-lijst;
- nagaan in hoeverre onderlinge verschillen in perceptie bestaan omtrent verheldering (clarificatie) en interpretatie, de twee sleutelbegrippen in de RFE-C.

Ruim een maand voor de start van de

registratieperiode is aan deze onderwerpen aandacht besteed. Wekelijks kwam de groep gedurende ten minste twee uren bijeen. In de eerste week is aandacht besteed aan het eerste punt. De overige vier weken kwamen ook de resterende drie punten aan bod. Gedurende deze weken hebben de huisartsen variabele aantallen redenen van contact opgeschreven en de daarbij behorende codes vastgelegd.

Tijdens de bijeenkomsten zijn ervaringen uitgewisseld en knelpunten besproken. Onvolkomenheden in de classificatie- en instructielijst werden vastgelegd. Deze werden voor de definitieve start

Betrouwbaarheid, adequaatheid en hanteerbaarheid

De redenen waarom mensen contact met de huisarts zoeken, vormen door hun veelheid en diversiteit een onoverzichtelijke massa van gegevens. Hierin is alleen orde te scheppen door ze te rangschikken. Wanneer dit op systematische wijze gebeurt, spreken we van classificeren of meten.⁸ Onder classificeren verstaan we het indelen van kenobjecten (in ons geval: redenen voor contact) in gelijksoortige klassen. De matrices gevormd door hoofdstukken en componenten zijn zulke klassen. Hierbinnen zijn weer codes te onderscheiden: hieraan kunnen de redenen van contact gekoppeld worden. Elke klasse, maar ook elke code daarbinnen, is omschreven, van een benaming voorzien en door middel van een nominale definitie omschreven.

Voorbeeld van een klasse: hoofdstuk *social problems* (verticaal) en component 1: *symptoms and complaints* (horizontaal).

Voorbeeld van een code binnen deze klasse: Z 10 *conjugal/marital problems*.

De klacht „ik heb de laatste tijd steeds ruzie met mijn man”, zou als code Z 10 gescoord worden.

Bij classificatie gaat het derhalve om het plaatsen van waargenomen objecten (de redenen van contact) in een ordeningssysteem. De betekenis hiervan wordt ontleend aan de waarde die huisartsen aan dit systeem toedelen. Hiermee komen we op het begrip adequaatheid.

De RFE-classificatielijst is *adequaat* wanneer een forum van huisartsen op basis van ervaringen met deze lijst aangeeft dat:

- de indeling of het geheel van classificaties aansluit op het huisartsgeneeskundig referentiekader;
- de indeling of het geheel van classificaties het totale scala van redenen van contact omvat; anders geformuleerd: wanneer alle redenen van contact zoals deze door huisartsen worden gepercipieerd – op het niveau van clarificatie of verheldering – in een van de codes van de RFE-lijst kunnen worden ondergebracht.

Adequaatheid impliceert geenszins betrouwbaarheid; niettemin zal toename van adequaatheid de betrouwbaarheid positief beïnvloeden.

Voor de volledigheid wijzen we erop dat we de term geldigheid of validiteit in het bovenstaande bewust hebben vermeden. Een geldigheids- of validiteitsmeting van de classificatielijst is in strikte zin niet mogelijk. In plaats hiervan hebben we de term *adequaatheid* ingevoerd. Bij geldigheid gaat het er om of bepaalde begrippen die gebruikt worden om een deel van de „werkelijkheid” systematisch te beschrijven (theorie), ook inderdaad door bepaalde indicatoren en vervolgens variabelen worden gedekt. Met behulp van de RFE-C lijst pogen we wél de werkelijkheid zoals huisartsen deze percipiëren te inventariseren of te beschrijven, maar de gebruikte RFE-C klassen en codes vormen geenszins indicatoren (en variabelen) van een theoretisch kader als systematisch geheel van samenhangende begrippen. Methodologisch bezien verwijzen indicatoren via een operationalisering ervan in variabelen immers naar een empirisch verschijnsel dat als representant beschouwd kan worden van een theoretisch begrip als onderdeel van een breder theoretisch kader. Geldigheid ligt als het ware in het verlengde van adequaatheid. We zouden kunnen zeggen dat adequaatheid een impliciet of pre-concept van geldigheid is: de uitspraak of de RFE-C al dan niet adequaat is, berust op een belangrijk criterium, namelijk het min of meer impliciete huisartsgeneeskundige referentiekader.

In hoeverre een dergelijk kader door middel van onderzoek geëxpliciteerd kan worden in de richting van theoretische (huisartsgeneeskundige) concepten, is een boeiende vraag, maar deze valt buiten het kader van dit artikel.

De RFE lijst is *betrouwbaar* wanneer onafhankelijke waarnemers een zelfde stimulus, dit is eenzelfde klacht, probleem of

vraag, een zelfde code toedelen. Betrouwbaarheid impliceert derhalve overeenstemming. Twee belangrijke voorwaarden voor betrouwbaarheid zijn:

- wederzijdse uitsluiting van de gebruikte klassen en codes;
- een goede operationalisering: dit slaat in geval van registratie op het formuleren van eenduidige instructies die nauwkeurig preciseren hoe het betreffende verschijnsel – reden van contact – zal worden geregistreerd. Hiermee kunnen vooral onsystematische fouten voorkomen worden. Onsystematische fouten kunnen ook te maken hebben met onderzoekersgebonden kenmerken als vermoeidheid, ongemotiveerdheid, afgeleid zijn door afwezigheid van derden of onvoldoende bekendheid met de indeling van classificatielijst.

De RFE-lijst is *praktisch hanteerbaar*, wanneer de huisartsen in de praktijk gebruik kunnen maken van deze lijst en de daaraan ten grondslag liggende instructies; dat wil zeggen dat zij met de classificatie en de instructies snel en gemakkelijk kunnen werken.

Concluderend kunnen we stellen dat onderzoeksresultaten beïnvloed kunnen worden door:

- de opzet en indeling van het onderzoeksinstrument;
 - de onderzoeksinstructies op operationalisering van het instrument;
 - houding en werkwijze van onderzoekers;
 - praktische hanteerbaarheid.
- Wanneer verschillende huisartsen bij dezelfde klacht onafhankelijk van elkaar verschillende codes toedelen, kan dit derhalve wijzen op onbetrouwbaarheid van de classificatielijst, maar ook op onbetrouwbaarheid van de instructies en een „onzorgvuldige” werkwijze van die huisartsen. Onzorgvuldigheid kan door geringe hanteerbaarheid in de hand gewerkt worden.

besproken met een lid van de WHO-werkgroep, ten einde zoveel mogelijk internationale afstemming te bewerkstelligen. Gedurende een aantal sessies werden voor het onderwerp beschikbare videobanden van consulten gepresenteerd om na te gaan of de huisartsen bij deze consulten verschillen vertoonden in het vastleggen van de formulering van de patiënt. Deze sessies bleken essentieel te zijn: de neergeschreven reden van contact moet immers het resultaat zijn van verheldering en uitdrukkelijk niet van interpretatie.

Registratieformulieren

De registratie vond plaats op speciaal door de WHO ontworpen dagformulieren.⁶ Per contact noteerden de huisartsen met behulp van codes de plaats van het contact (bijvoorbeeld praktijk of thuis), alsmede geslacht en leeftijd van de patiënt. Vervolgens schreven zij de reden van contact uit, voorafgegaan door de RFE-classificatie. De artsen noteerden op de achterkant van het formulier de eventuele problemen die de classificatie hen opleverde. Deze gegevens werden centraal verzameld en verwerkt.

Betrouwbaarheid of overeenstemming

Om uitspraken te kunnen doen over de betrouwbaarheid van de RFE-lijst hebben we twee aselechte steekproeven getrokken uit de tijdens de veldoefening verzamelde registratieformulieren. Deze steekproeven bevatten respectievelijk 1.690 en 1.740 redenen van contact. Elk pakket werd door vier huisartsen nogmaals gecodeerd. Per contactreden waren leeftijd en geslacht van de patiënt gegeven.

De overeenstemming tussen de coderende huisartsen is berekend met de kappa (k).⁹ Hiermee wordt aangegeven hoeveel groter de overeenstemming is dan volgens louter toeval verwacht zou mogen worden. Bij een kappa van bijvoorbeeld .80 is de overeenstemming 80 procent groter dan op grond van het toeval zou kunnen worden verwacht. De kappa-waarde kan variëren van 0 tot 1,00, waarbij 0 staat voor geen enkele overeenstemming en 1,00 voor volledige overeenstemming.

Inventarisatie foutenbronnen

Tijdens de wekelijkse bijeenkomsten werden knelpunten geïnventariseerd en zo nodig in (nieuwe) instructies vertaald. De conclusies hieromtrent zijn door onszelf vastgelegd en doorgespeeld naar de WHO-werkgroep. Tijdens de registratie kon elke arts ondervon-

den problemen vastleggen op de achterzijde van het RFE-registratieformulier. Deze „achterkant-problemen” werden uitgetypt en per arts gebundeld. Vervolgens analyseerden koppels van twee huisartsen de problemen van twee collega's, met behulp van de volgende categorieën:

1. Classificatielijst voorziet in het probleem; is geen (echt) probleem meer.
2. Instructielijst voorziet in het probleem; is geen (echt) probleem meer.
3. Afspraken die gemaakt zijn, onderwerpen het probleem; is geen (echt) probleem meer.
4. Blijft een probleem omdat...

Elke arts heeft daarna zijn eigen „probleembundel” met de opmerkingen van zijn collega's teruggekregen en deze problemen zelf nog eens bekeken. Voor zover men het niet eens was met de beoordelaars, is dit met hen besproken, ten einde tot een gezamenlijke oplossing te komen. Voor zover men elkaar toch niet kon vinden, werd het probleem alsnog aan de plenaire vergadering voorgelegd. In de vergadering werd een beslissing genomen omtrent de definitieve probleemformulering en analyse.

Nadat 800 redenen van contact waren verzameld, zijn deze redenen uitgetypt en aan de acht huisartsen voorgelegd, in die zin dat twee groepen van vier elk 400 contactredenen kregen voorgelegd. Zij hebben onafhankelijk van elkaar geëvalueerd. Bij een overeenstemming van 50 procent of lager werden de betreffende contactredenen en toegevoegde codes nader geanalyseerd op de foutenbron; ook dit werd gedaan door paren van twee huisartsen. Deze paren werd gevraagd aan te kruisen welke fout van toepassing was:

1. Er is een onvoldoende verhelderde klacht neergeschreven en vervolgens geëvalueerd.
2. De klacht is wel goed uitgeschreven, maar er is verkeerd geëvalueerd, zonder dat dit het gevolg was van fouten in de instructie- of de classificatielijst.
3. De instructie was niet toereikend. De fouten 1 en 2 hebben we aangeduid als arts-gebonden fouten; fout 3 is een lijst-gebonden fout.

Adequaatheid en praktische hanteerbaarheid

Tenslotte is gedurende de wekelijkse besprekingen ook regelmatig aandacht besteed aan de ervaringen met de RFE-C in termen van adequaatheid en praktische hanteerbaarheid.

Resultaten

Betrouwbaarheid

Tabel 1 geeft een overzicht van de betrouwbaarheid van de RFE-C. Gelet op de totaalscores – .87 en .91 – mag gesproken worden van een hoge betrouwbaarheid.

Bezien we de pakketten A en B, dan valt op dat de spreiding in pakket A varieert van .69 tot .96, en in pakket B van .61 tot .96. Niettemin liggen 29 van de 34 scores op .80 en hoger. Tevens is vermeldenswaard dat in de hoofdstukken F, H en R de hoogste overeenstemmingsscores gevonden worden, terwijl hoofdstuk Y de laagste waarde heeft.

Wanneer we de scores van de twee steekproeven vergelijken, valt op dat ze elkaar niet opvallend ontlopen, behalve in de hoofdstukken B en Y, waar de verschillen iets groter zijn.

In tabel 2 zien we eveneens hoge overeenstemmingen, respectievelijk .78 en .79, zij het dat de scores in vergelijking met tabel 1 wat lager zijn. Gelet op het strenge karakter van de kappa, mogen we niettemin ook hier spreken van een hoge betrouwbaarheid. In de afzonderlijke componenten valt op dat 8 van de 14 scores op .80 en hoger liggen.

De scores van A en B ontlopen elkaar ook hier nauwelijks; component 6 vormt hierop echter een uitzondering. Deze component blijkt bij beide groepen (A én B) in vergelijking met de andere componenten het laagst te zijn. Kennelijk geeft *other* als restcategorie hoe dan ook interpretatieproblemen. De enige opmerking die we over het verschil tussen A en B kunnen maken, is dat de artsen-codeurs van groep B kennelijk beter in staat zijn geweest de uitgeschreven reden voor contact te classificeren. Deze interpretatie berust op de constatering dat de totaalscores van B in beide tabellen iets hoger zijn dan die van A.

Inventarisatie foutenbronnen

De vastgestelde probleemformuleringen, de analyse ervan en de daaruit voortvloeiende adviezen ter bijstelling van de classificatie en instructielijst laten we in het kader van dit artikel buiten beschouwing, temeer omdat deze in de bijgestelde RFE-lijst – c.q. de ICPC – ondervangen zijn. Vermeldenswaard achten we echter de resultaten van de analyse van de score van lage overeenstemming: bij 148 van de 800 redenen van contact (18 procent) was de mate van overeenstemming 50 procent of lager.

Uit tabel 3 blijkt dat de meerderheid van de fouten niet lijst-, maar artsgebonden is. Van de gemaakte fouten blijkt naar het oordeel van de deelnemende artsen 79 procent (44 en 35 procent) artsgebonden te zijn. De voornaamste fout die gemaakt werd, was het onvoldoende duidelijk neerschrijven van de reden van contact; het onvoldoende opvolgen van de instructies komt op de tweede plaats. Deze artsgebonden fouten kunnen opgevangen worden door betere instructies en training.

In slechts 21 procent van de gevallen is er sprake van fouten die opgevangen kunnen worden door uitdrukkelijke aanpassing van de onderscheiden lijsten (fout 3).

Praktische hanteerbaarheid en adequaatheid

Het registreren en classificeren van de redenen van contact van patiënten met hun arts was in de beginfase tijdrovend. Na een inwerkperiode van enkele weken werd het omgaan met de classificatie- en instructielijst gemakkelijker. Per consult werd in het begin van de registratie gemiddeld twee minuten gebruikt om de reden van contact te coderen, later gemiddeld één minuut; de variatie lag toen tussen 15 seconden en twee minuten.

De benodigde tijd betrof soms het vaststellen van de code, vooral wanneer er sprake was van afgrenzing ten opzichte van aangrenzende codes. Soms was, mede in verband met het voorgaande, extra tijd nodig voor reflectie over de contactredenen en voor eventueel navragen bij de patiënt van de reden van contact. In veel gevallen was de reden van contact en de daarbij behorende code echter evident. Een kwart minuut extra per reden van contact bleek op den duur voor vrijwel iedereen toereikend. Deze tijd was voornamelijk nodig voor het verkort uitschrijven van de contactredenen. Gelet op het *feasibility*- of haalbaarheidsaspect was deze extra tijd alleszins acceptabel. De verwachting van de huisartsen was dat de extra tijdsbesteding gelijk blijft en mogelijk nog iets verder zal afnemen wanneer de lijst en instructies zullen zijn bijgesteld.

Ondanks de goede RFE-handleiding, was er nogal eens behoefte aan uitbreiding van instructies, aan een verscherpte afbakening van de inhoud van de verschillende classificaties en soms aan toevoeging van synoniemen. Deze behoefte kwam vooral naar voren tijdens de regelmatige besprekingen. Deze waren vooral in de beginfase essentieel: ervaringen en problemen werden be-

sproken terwijl op basis hiervan nadere afspraken omtrent coderingen werden gemaakt. Vooral de hoofdstukken A, L, P en S werden veel besproken, met name de eerste component daarvan:

- In hoofdstuk A leek vaak een dubbelcodering mogelijk, doordat onvoldoende werd gezocht naar meer specifieke coderingsmogelijkheden.
- Hoofdstuk L was vooral problematisch bij de overgangsgebieden tussen romp en extremiteiten, respectievelijk romp en hoofd.

• De codes van component 1 van hoofdstuk S waren vaak onderling moeilijk af te grenzen.

• Component 1 van hoofdstuk P werd vaak besproken in relatie tot component 7 van hetzelfde hoofdstuk en tot hoofdstuk Z.

Het coderen van herhalingsconsulten vroeg om duidelijke afspraken, omdat daarin gemeenschappelijke voorkennis van patiënt en arts vaak niet meer gevaloriseerd wordt.

Het gebruik van diagnoses door de pa-

Tabel 1 – Table 1. De betrouwbaarheid (kappa-waarden) van de RFE-C hoofdstukken – The reliability (kappa-values) of the RFE-C chapters.

Chapters	A (n=1690)	B (n=1740)
A. General and unspecified	.70	.80
B. Blood, -forming, organs, lymphatics, pleen	.69	.86
D. Digestive	.82	.91
F. Eye	.96	.96
H. Ear	.96	.95
K. Circulatory	.92	.94
L. Musculoskeletal	.91	.93
N. Neurological	.90	.94
P. Psychological	.93	.90
R. Respiratory	.96	.96
S. Skin	.89	.92
T. Endocrine, metabolic and nutritional	.88	.87
U. Urology	.89	.91
X. Female genital system	.81	.93
Y. Male genital system	.70	.61
Z. Social problems	.82	.75
Total	.87	.91

Tabel 2 – Table 2. De betrouwbaarheid (kappa-waarden) van de RFE-C componenten – The reliability (kappa-values) of the RFE-C components.

Components	A (n=1690)	B (n=1740)
1. Symptoms and complaints	.86	.85
2. Diagnostic, screening and preventive procedures	.69	.72
3. Treatment procedures and medication	.81	.84
4. Results	.89	.96
5. Administrative	.83	.85
6. Other	.68	.34
7. Diagnosis/diseases	.65	.70
Total	.78	.79

Tabel 3 – Table 3. Oorzaken lage scores – Causes of low scores.

	Number	Percentage
Reason for encounter not fully written out	63	44
Wrong code due to coder error	51	35
Wrong code due to deficiency of instruction and classification list	34	21

tiënt was eveneens vaak aanleiding tot discussie.

Een probleem van andere orde was het werken met op papier uitgeschreven redenen van contact. Het opnieuw coderen van uitgeschreven klachten gaf nogal eens aanleiding tot verschillende coderingen. Oorzaak was soms dat de tekst onvoldoende weergaf wat er tijdens het consult aan de orde was geweest, soms dat de arts in zijn omschrijving reeds had geïnterpreteerd. Dit laatste wijst erop dat het verhelderen, zonder daarin reeds tot interpretatie over te

gaan, steeds een punt van aandacht moet zijn.

Het classificeren aan de hand van papieren redenen van contact was niet alleen voor de onderzoekers van belang om tot een eerste globale betrouwbaarheids-schatting te komen, maar vooral voor de deelnemende huisartsen om tot een duidelijke afbakening te komen van de begrippen verheldering en interpretatie.

Er bleken nauwelijks problemen op te treden met betrekking tot de adequaat-

heid van de RFE-classificatie: de indeling van de lijst in hoofdstukken en componenten, alsmede de onderverdeling in codes, sluit in het algemeen goed aan op de (geneeskundige) referentiekaders van de huisartsen. Enkele huisartsen gaven aan dat de hoofdstukken P en Z te weinig specifiek zijn en daarom minder adequaat. Men kon echter niet tot overeenstemming komen inzake de gewenste bijstellingen. Vooral achtte men P en Z acceptabel, hoewel een nader onderzoek gewenst geacht wordt.

Synopsis

Van der Horst F, Seelen A, Vissers F, Plagge H, Von Höfen R, De Geus C, Kleijnen J. Registration in general practice. Reliability, feasibility and adequacy of the RFE classification system. Huisarts en Wetenschap 1985; 28: 229-34.

Introduction. The aim of a research group of the University of Limburg is to establish a registration network for family practices to provide a systematic and coordinated description of input (S), transit (O and A) and output (P) of problems presented to the general practitioner. This article reports on our experience with registration of the S (subjective). The problem definition of the study was: to what extent is the Reason for Encounter Classification (RFE-C) a reliable and, from the point of view of the general practitioner, feasible and adequate instrument for classification of the input (S) in terms of reasons for encounter.

Methods. Detailed reports on the RFE-C have been published in previous issues of *Huisarts en Wetenschap*.⁷

Participants in the study were eight general practitioners divided over one solo practice (one doctor), one association practice (two doctors) and three health centres (six doctors). The total practice population of the participants was about 22,000 patients. The field trial took three months and covered 10,046 patient encounters and 15,195 reasons for encounter.

Registration was done on day forms designed by the WHO. The general practitioners used codes to register the site of encounter (e.g. home or practice) and the patient's age and sex per encounter. Next they wrote out the reason for encounter, preceded by the RFE code. Problems, if any, were noted on the reverse side of the form. These data were centrally collected and processed.

In order to permit conclusions about the reliability of the RFE list, two random samples were drawn from the registration forms collected during the field trial. These samples comprised 1,690 and 1,740 reasons

for encounter respectively. Each package was coded by four general practitioners, giving the patient's age and sex per reason for encounter.

Concurrence between the coding general practitioners was calculated with the aid of the kappa-value (K)⁹ which indicates by how much the concurrence exceeds that to be expected on the basis of pure chance. A kappa-value of, say, .80 indicates a concurrence which is 80 percent in excess of that to be expected on the basis of pure chance. The kappa-value can range from 0 (no concurrence at all) to 1.00 (total concurrence).

Results. Table 1 presents data on the reliability of the RFE-C. In view of the total scores – .87 and .91 – this can be described as high. A look at packages A and B reveals that the range in package A is from .69 to .96, while that in package B is from .61 to .96. Nevertheless, 29 of the 34 scores are .80 and over. Comparing the scores of the two random samples we find that there are no striking differences except in chapters B and Y.

Table 2 likewise shows high degrees of concurrence (.78 and .79 respectively), albeit that the scores are slightly lower than those in table 1. Given the strictness of K, however, reliability can again be described as good. In the separate components it is conspicuous that eight of the fourteen scores are .80 or over. Again, the scores of A and B hardly differ, with the exception of component 6. It is evident that the rest category *other* somehow poses problems of interpretation.

Table 3 shows that most errors are not list-but doctor-related. It should be possible to eliminate these errors by improved instruction and training.

Registration and classification of the reason for encounter was time-consuming in the initial phase of the study. After a working-in period of a few weeks the use of the instruction and classification list became easier. In the initial phase the coding of the reason for encounter took an average of two minutes per encounter; later on, 15 seconds proved to be sufficient for virtually all participants. This time was spent mainly

on a brief write-out of the reason for encounter. In view of the *feasibility* aspect, this additional time is quite acceptable.

There were hardly any problems regarding the *adequacy* of the RFE-C: the division of the list into chapters and components and the subdivision into codes are generally compatible with the (medical) reference frames of the participating general practitioners.

Discussion. The findings warrant the conclusion that the RFE-C is a reliable and adequate instrument for routine use in general practice. Moreover, registration on the basis of adjusted lists can be expected to lead to higher reliability scores (above .90) and less loss of time (less than 15 seconds per reason for encounter). In this respect it is to be borne in mind, however, that the satisfactory results of this study are partly due to the commitment and critical attitude of the participating general practitioners. We used a version of the classification list which is no longer in use. This list has been adjusted on the basis of our experience and that of general practitioners in other countries. Moreover, the RFE-C is to be integrated in a new system: the International Classification of Primary Care (ICPC). This list is based on the RFE list but has so been adjusted that three of the four SOAP elements – S, A, and P – can be fully classified with it.

Uniform registration and classification with the aid of a system that is reliable, adequate and feasible for the general practitioner opens up new perspectives for inter-peer auditing and consultation, further research, and teaching. In any case the ICPC provides a structure for the empirical description of the problem-solving process in general practice, and thus opens new roads to a further evolution of general practice.

Keywords: Classification; Family practice; Reason for encounter; Registration.

Address of authors: Capaciteitsgroep Huisartsgeneeskunde, Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht, The Netherlands.

Beschouwing

Het registreren en classificeren voor dit onderzoek is verricht door full-time werkende huisartsen. Deze huisartsen zijn hiervoor bewust aangezocht, omdat wij een beeld wilden krijgen van het spanningsveld tussen hetgeen noodzakelijk dan wel wenselijk is vanuit een zuiver wetenschappelijk gezichtspunt, en hetgeen haalbaar is in de hitte van de dagelijkse praktijk.

Uit het onderzoek is duidelijk geworden, dat de RFE-classificatie een betrouwbaar en adequaat instrument is, waarmee in de dagelijkse praktijkvoering goed kan worden gewerkt. Verwacht mag bovendien worden dat het registreren op basis van bijgestelde lijsten zal leiden tot hogere betrouwbaarheidsscores (boven .90) en minder tijdverlies (minder dan 15 seconden per contactreden). Daarbij dient overigens aangetekend te worden dat de bevredigende uitkomsten van dit onderzoek mede zijn te danken aan de inzet en kritische houding van de deelnemende huisartsen.

De door ons gehanteerde versie van de classificatielijst wordt thans niet meer gebruikt. Deze lijst is bijgesteld, op basis van onze ervaringen en die van huisartsen uit andere landen. De RFE-C is opgegaan in een nieuw systeem, de International Classification of Primary Care (ICPC). Deze lijst is weliswaar geënt op de RFE-lijst, maar zodanig bijgesteld, dat hiermee drie van de vier SOEP-elementen – S, E en P – volledig geclassificeerd kunnen worden.

We mogen aannemen dat bij een continue registratie van deze elementen de extra benodigde tijd niet meer dan één minuut per reden voor contact zal bedragen. Wanneer daarnaast ook het lichamelijke onderzoek (O) zal worden vastgelegd met behulp van een aparte lijst, zal deze termijn waarschijnlijk

worden overschreden. Toch mogen we nauwelijks spreken van „extra tijd”, omdat een goede klassieke verslaglegging immers ook tijd kost.

Uniforme registratie en classificatie met behulp van een voor de huisarts betrouwbaar, adequaat en praktisch hanteerbaar systeem biedt uitdrukkelijke extra perspectieven voor intercollegiale toetsing en overleg, verder wetenschappelijk onderzoek én onderwijs. Met de ICPC zal het probleemoplossend proces van de huisarts in ieder geval adequaat empirisch onderbouwd kunnen worden. Hiermee worden nieuwe wegen geopend voor een verdere ontwikkeling van huisartsgeneeskunde.

Conclusies

Gelet op de resultaten, komen we met betrekking tot de in de inleiding vermelde vraagstelling tot de volgende conclusies:

- de RFE-classificatie is een adequaat ordeningssysteem, in die zin dat het:
 - aansluit op de geneeskundige referentiekaders van de huisarts;
 - de totale *range* van de door de huisarts gespecificeerde redenen van contact omvat.

De hoofdstukken P en Z verdienen nadere uitwerking in toekomstig onderzoek.

- De betrouwbaarheid van de RFE-classificatie kan op basis van de analyse-resultaten als goed tot zeer goed betiteld worden. Daarbij dient aangetekend te worden dat de betrouwbaarheid op basis van de componenten-analyse lager ligt dan die op basis van de hoofdstukken-analyse.

- Gelet op de *feasibility* blijkt een RFE-classificatie in de praktijk mogelijk te zijn. De extra tijdsbelasting blijkt in het begin twee minuten per reden van contact te zijn, later dalend tot een kwart minuut.

Dankbetuiging

Wij zijn Dr. J. Diederiks en J. Joosten – medisch sociologen – en Prof. dr. H. Lamberts, huisarts en lid van de WHO-werkgroep International Classification for Primary Care (ICPC) erkentelijk voor de kritische wijze waarop zij het concept hebben becommentarieerd. Wij zijn hen zeer erkentelijk voor de gegeven adviezen.

¹ Van der Horst F, Stalenhoef P, Vierhout W. Probleemgeoriënteerde verslaglegging in de huisartspraktijk. 1. Ervaringen. Huisarts en Wetenschap 1981; 24: 252-8.

² De Geus CA, Van Höfen R, Metsemakers J. Huisarts en computer. Med Contact 1984; 40: 1289-90.

³ Meads S. Reason for encounter classification. Field Trial Manual. WHO Chronicle 1983; 37, 159-62.

⁴ Lamberts H, Meads S, Wood M. The reason for encounter classification for use in primary care. A pilot study in the Netherlands. In: Van Bommel JH, Ball MJ, Wiggert O, eds. Medinfo-83. Proceedings of the Fourth World Conference on Medical Informatics. Amsterdam: North Hilland Publishing Company, 1983.

⁵ De Geus CA, Van der Horst F. Lage rugklachten als reden om de huisarts te raadplegen. Huisarts en Wetenschap 1983; 26, (suppl Huisarts & Praktijk): 28-31.

⁶ Lamberts H, Meads S, Wood M. Classification of Reasons why persons seek primary care. Pilot-study of a new system. Public Health Rep 1984; 99: 597-605.

⁷ Lamberts H, Meads S, Wood M. Waarom gaat iemand naar de huisarts? Een internationale studie met de Reason for Encounter Classification. Huisarts en Wetenschap 1984; 27: 234-44.

⁸ Gadourek I. Sociologische onderzoeksmethoden. Deventer: Van Loghum Slaterus, 1972.

⁹ Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968; 70: 213.