

Het ontwikkelen van classificatiesystemen bij aandoeningen van het bewegingsapparaat

SMA BIERMA-ZEINSTRASMA, AP VERHAGEN, MY BERGER

Inleiding

Al eeuwen heeft men in de geneeskunde getracht de verschillende aandoeningen onder te brengen in een structuur, zodat men deze aandoeningen in de praktijk gemakkelijker en meer gestandaardiseerd kan herkennen. Zowel Sydenham (17e eeuw) als De Sauvage (18e eeuw) hebben getracht gedetailleerde classificatiesystemen te ontwerpen op grond van een symptomatologie waarin alle aandoeningen konden worden ondergebracht.¹ Naarmate in de 19e eeuw de medische technologie zich ontwikkelde, werden de aandoeningen steeds meer ingedeeld op grond van structurele of fysiologische verschijnselen. Binnen de huisartsgeneeskunde is de ICPC waarschijnlijk de meest bekende classificatie. Daarnaast zijn in diverse standaarden – bijvoorbeeld de NHG-Standaard Schouderklachten – classificaties geïntroduceerd.

Een classificatie kan een indeling naar diagnose beogen, maar kan ook een grovere indeling geven, waarin groepen diagnoses met voldoende overeenkomsten voor eenzelfde beleid worden samengevoegd tot één klasse; dat gebeurt bijvoorbeeld in de NHG-Standaard Schouderklachten. Ook kan een classificatie een indeling naar stadia van ziekte beogen of een indeling naar functionele klachten of psychosociale risicofactoren.

Bij een classificatie hoort in beginsel een gestandaardiseerde beschrijving van de te onderscheiden groepen, en eventueel criteria op grond waarvan deze groepen kunnen worden onderscheiden. Het doel van deze criteria zal van invloed zijn op hun samenstelling. Is het doel het stellen van een diagnose bij een individuele patiënt, dan zal de waarde die men hecht aan een 'foute classificatie' afhankelijk zijn van die diagnose. Bij de overweging een patiënt te verwijzen met verdenking op maagcarcinoom zal men het niet-missen van de diagnose (fout-negatief) belangrijker vinden dan het ten onrechte verwijzen van de patiënt (fout-positief). Hier zal men een fout-positieve classificatie graag accepteren wanneer dit tot minder fout-negatieve classificaties leidt. Wanneer

Samenvatting

Bierma-Zeinstra SMA, Verhagen AP, Berger MY. Het ontwikkelen van classificatiesystemen bij aandoeningen van het bewegingsapparaat. Huisarts Wet 2000;43(11):473-7.

Veel aandoeningen waarmee de huisarts regelmatig te maken krijgt, zijn syndromen. Het stellen van een diagnose is vaak moeilijk bij dergelijke aandoeningen, omdat een gouden standaard ontbreekt. Dit maakt het ontwikkelen van criteria ter classificatie van dergelijke aandoeningen lastig. Meestal worden deze classificaties op ongestructureerde en niet-reproduceerbare wijze samengesteld. In deze beschouwing bespreken wij enkele gestructureerde methoden om een classificatiesysteem te ontwikkelen. Eén van deze methoden is de clusteranalyse; deze aanpak wordt aan de hand van een praktijkvoorbeeld verder uitgelegd. Voordat een classificatiesysteem uiteindelijk in de dagelijkse praktijk van de huisarts en/of de onderzoeker kan worden geïntroduceerd, moet het worden beoordeeld op validiteit, reproduceerbaarheid en bruikbaarheid. Daarnaast moet altijd nagegaan worden of het classificatiesysteem wel is ontwikkeld voor het doel (bijvoorbeeld verslaglegging of gezondheidszorg-onderzoek) waarvoor het gebruikt zal worden.

Instituut Huisartsgeneeskunde, Erasmus Universiteit Rotterdam, Postbus 1738, 3000 DR Rotterdam.
dr. S.M.A. Bierma-Zeinstra, gezondheids-wetenschapper; dr. A.P. Verhagen, gezondheidswetenschapper-epidemioloog; dr. M.Y. Berger, huisarts-epidemioloog.
Correspondentie: dr. S.M.A. Bierma-Zeinstra; e-mail: bierma@hag.fgg.eur.nl

men echter classificeert met het doel vast te stellen hoeveel maagcarcinomen jaarlijks in Nederland voorkomen, speelt dit onderscheid nauwelijks een rol; hier gaat het om de accuratesse waarmee men meet (zo weinig mogelijk misclassificaties).

Het onderscheidend vermogen van dergelijke criteria test men het liefst aan de hand van een gouden standaard, die een onomstotelijk bewijs levert voor de aanwezigheid van een aandoening, bijvoorbeeld een biopsie of een scan. Voor slechts weinig aandoeningen in de huisartspraktijk is een dergelijke objectieveerbare gouden standaard beschikbaar. Veel van de in de huisartsgeneeskunde voorkomende aandoeningen zijn syndromen (een patroon van symptomen dat herhaaldelijk optreedt bij verschillende patiënten), zoals depressie en aandoeningen van het bewegingsapparaat. De classificatie van dergelijke aandoeningen is een stuk lastiger.²

In deze bijdrage bespreken wij hoe dergelijke classificaties worden gemaakt en welke methodologische valkuilen men daarbij tegenkomt.

Methoden voor het ontwerpen van een classificatie

Classificaties van klinische syndromen worden meestal opgesteld op grond van ideeën van de ontwerpers die weer gebaseerd zijn op medische literatuur en/of eigen ervaring, al dan niet aangevuld met een consensusprocedure. Vaak wordt alleen een classificatie gemaakt met een beschrijving van de symptomatologie van de verschillende groepen patiënten, soms zijn er ook classificatiecriteria.

Er zijn enkele gestructureerde methoden om een op symptomen gebaseerde classificatie en vervolgens classificatiecriteria te ontwerpen. Het American College of Rheumatology (ACR) gebruikt een dergelijke methode om tot classificatiecriteria van reumatische aandoeningen te komen.³ Hierbij classificeert men steeds één aandoening ten opzichte van de andere aandoeningen. Via een schriftelijke, anonieme consensusmethode (Delphi-methode) selecteren experts de voor de desbetreffende aandoening belangrijkste symptomen. Het voorkomen van deze symptomen wordt daarna prospectief onderzocht in een relevante populatie. De arts die de patiënt insluit in het onderzoek, beoordeelt of deze de specifieke aandoening

heeft, en noteert de aanwezigheid van de geselecteerde symptomen. Een team van experts controleert de 'juistheid' van de diagnose op basis van aan- of afwezigheid van de genoteerde symptomen. Hierna bepaalt men voor de gehele onderzoekspopulatie welke symptomen de hoogste specificiteit en sensitiviteit hebben ten aanzien van de aanwezigheid van de aandoening. Vervolgens kunnen met behulp van statistische analyses (discriminantanalyse, logistische regressie-analyse, de 'classification and regression tree'-methode) classificatiecriteria worden benoemd.

Een geheel andere benadering is de clusteranalyse.⁴ In het mathematisch model bij clusteranalyse bevinden de patiënten zich in een multi-dimensionele ruimte met net zoveel assen als er symptomen (variabelen) zijn. In deze ruimte wordt bepaald hoe veraf of dichtbij de patiënten zich van elkaar bevinden. Patiënten die op alle symptomen *gelijk* scoren, bevinden zich op hetzelfde punt in de ruimte. Patiënten die op slechts enkele symptomen verschillen, zullen zich in de ruimte dicht bij elkaar bevinden. Er kunnen nu groepen worden onderscheiden op grond van 'gelijkheid' in symptomatologie en men kan het patroon van symptomen voor de verschillende groepen beschrijven. Zo hoopt men dat de gecreëerde groepen patiënten de verschillende syndromen (of verschillende stadia van syndromen) representeren. Ook hier kan men vervolgens weer met statistische analyses classificatiecriteria benoemen.

Een voorbeeld uit de praktijk

Clusteranalyse is gebruikt in twee Nederlandse onderzoeken voor het opstellen van een classificatie van schouderaandoeningen in de huisartspraktijk.^{5,6} Daarnaast is clusteranalyse recent gebruikt voor het opstellen van een classificatie van heup-aandoeningen bij ouderen in de huisartspraktijk.⁷

De procedure in dat laatste onderzoek is illustratief. Patiënten (n=224) van 50 jaar en ouder die de huisarts hadden geconsulteerd wegens heupklachten en werden

De kern

- Bij elk classificatiesysteem moet men zich afvragen of het is gemaakt voor het doel waarvoor men het wil gebruiken.
- Bij elk classificatiesysteem moet men zich afvragen of is onderzocht of het betrouwbaar de werkelijkheid beschrijft (validiteit).
- Classificatiecriteria kunnen verschillen per setting.
- Classificatiecriteria kunnen verschillen per doel (bijvoorbeeld individuele patiëntenzorg versus epidemiologisch onderzoek).

verwezen voor röntgenonderzoek van de heup, ondergingen een gestandaardiseerde anamnese en lichamelijk onderzoek van lage rug, heup en knie. Vervolgens werd een echografisch onderzoek uitgevoerd en werd de BSE bepaald. Clusteranalyse met gegevens uit anamnese en lichamelijk onderzoek van de heup (65 variabelen) resulteerde in een classificatie met negen symptomatologische clusters.

Een groep van 20 experts herkende, zonder de beschikking te hebben over de radiologische en sonografische gegevens, de combinatie van symptomen in zeven van de negen clusters als bekende syndromen uit de klinische praktijk. De combinaties van symptomen binnen deze zeven clusters toonden tevens gelijkenis met de symptomatologie van bekende en beschreven syndromen. De twee andere clusters waren moeilijk te identificeren; verder onderzoek zal moeten uitwijzen welke aandoeningen hieraan ten grondslag liggen.

Twintig symptomen waren zodanig discriminerend dat daarmee 80 procent van de patiënten nog steeds in dezelfde cluster

werd geplaatst als bij de oorspronkelijke procedure het geval was. De helft van deze twintig variabelen bestond uit variabelen uit de anamnese, en de exacte locatie van de pijn bleek hiervan het meest discriminerend. Van de bewegingsbeperkingen was niet de endorotatie maar de abductie de meest discriminerende variabele.

Methodologische problemen

Een probleem bij een door experts en/of consensus opgestelde classificatie en de bijbehorende classificatiecriteria is de cirkelredenering. Dit probleem doet zich ook voor bij de meer gestructureerde methode die gebruikt wordt door het American College of Rheumatology. Zo beschouwden de experts bij het opstellen van de ACR-classificatiecriteria voor artrose van de heup⁸ de endorotatiebeperking als een belangrijk symptoom. Daardoor werd dit symptoom bij iedere patiënt gemeten. Wanneer nu de aanwezigheid van een beperkte endorotatie ook een belangrijke factor is voor het besluit (door de clinicus en/of de experts) dat er sprake is van artrose van de heup, is het logisch dat dit symptoom bij de statistische analyses zal worden geselecteerd als één van de belangrijkste classificatiecriteria. Er is in een dergelijk geval echter geen nieuwe kennis gegenereerd, maar alleen op een nette manier consensus bereikt over iets wat we al dachten te weten.

Felson & Anderson⁹ stellen dat deze cirkelredenering is te voorkomen door verschillende experts te gebruiken voor het selecteren van de symptomen en voor het 'classificeren' van de patiënten. Het is echter de vraag of dit veel uitmaakt, omdat de meeste experts zullen uitgaan van dezelfde voorkennis.

Hoewel het classificeren met behulp van clusteranalyses in eerste instantie een objectieve methode lijkt, blijken ook hier veel methodologische valkuilen te bestaan. Allereerst zijn er zeer uiteenlopende clustermethoden die alle berusten op een andere rekentechniek om te bepalen welke patiënten het meest op elkaar lijken, met

als resultaat verschillende classificaties.⁴ Er bestaat geen gestandaardiseerd protocol voor het kiezen van de beste techniek bij een specifiek classificatieprobleem. De keuze moet worden gemaakt op basis van gezond verstand en men kan hoogstens vermelden waarom de keuze is gemaakt.

Vanzelfsprekend is de keuze van de variabelen die gebruikt worden voor de classificatie van groot belang. De variabelen moeten relevant zijn voor het doel van de classificatie. Er moet een plausibele relatie verondersteld kunnen worden tussen de variabele en de onderliggende aandoeningen. Wanneer men een indeling nastreeft van heupaandoeningen, zal men bijvoorbeeld niet de variabele 'haarkleur' meenemen. Om een zo objectief mogelijke classificatie te verkrijgen werden bij de classificatie van heupklachten alle variabelen van anamnese en lichamelijk onderzoek van de heupregio ingesloten, in plaats van alleen de op voorhand als belangrijk veronderstelde symptomen.

Het volgende besluit betreft het aantal clusters. Hierbij gaat het om een afweging van het klinische detail tegen de reproduceerbaarheid van de afzonderlijke clusters. Hoe kleiner de clusters, des te meer toevalligheidsaspecten een rol spelen. Er zijn verschillende technieken om het optimale aantal clusters te bepalen. *Milligan & Cooper*¹⁰ adviseren bijvoorbeeld om het

aantal clusters niet groter te maken dan de wortel uit het aantal variabelen of de wortel uit het aantal patiënten. In het geval van de classificatie van heupklachten zou het aantal clusters bij gebruik van 65 variabelen dus maximaal acht mogen zijn. Toch werd gekozen voor de oplossing met negen clusters om een verlies aan klinisch detail tegen te gaan.

Validiteit, reproduceerbaarheid, bruikbaarheid

Naast het feit dat er veel subjectieve keuzen gemaakt moeten worden, is clusteranalyse een zuiver mathematische benadering en ontbreekt een statistische toetsing. Hierdoor kan men systematische en toevalligheidsaspecten niet van elkaar onderscheiden. Men zal daarom de reproduceerbaarheid van een dergelijke classificatie goed moeten verifiëren. Deze reproduceerbaarheid is naast de validiteit een belangrijk onderdeel van de betrouwbaarheid van een classificatie. Voor de validiteit van een dergelijke classificatie zal men op een of andere manier moeten nagaan of de gevormde groepen (clusters) werkelijk syndromen lijken te vertegenwoordigen.

Hieronder staat een overzicht van de be-

langrijkste aspecten waarop men een classificatie zal moeten beoordelen.

- Bij het toetsen van de *constructvaliditeit* wordt vaak nagegaan of de in de classificatie te onderscheiden groepen een relatie vertonen met aspecten/factoren die passen bij de verschillende aandoeningen. In de classificatie van heupklachten werden positieve relaties gevonden tussen bepaalde clusters en radiologische en echografische verschijnselen die ook theoretisch te verwachten waren.

- Voor de *contentvaliditeit* is het belangrijk dat men de relevante patiëntgroepen en de relevante variabelen (die dus alle aspecten van de aandoening beschrijven) heeft ingesloten. In de classificatie van heupklachten waren inclusie- en exclusiecriteria gekozen met het doel de relevante patiëntengroep te selecteren: patiënten van middelbare en oudere leeftijd die de huisarts consulteren met een recent begonnen heuppijn (die eventueel chronisch kan worden). Alle symptomen uit de anamnese en het lichamelijk onderzoek werden ingesloten. De classificatie had aan contentvaliditeit gewonnen als ook de variabelen van het echografisch en radiologisch onderzoek waren ingesloten; deze werden echter gebruikt om de constructvaliditeit te toetsen.

Contentvaliditeit vergt bovendien dat de categorieën in een classificatie elkaar

Aspecten waarop classificaties (en hierop gebaseerde criteria) voor introductie moeten worden gescreend (vrij naar Felson & Andersson⁹ en Buchbinder et al.¹¹)

Doel • Is het doel van de classificatie of van de hierop gebaseerde criteria duidelijk aangegeven?	Bruikbaarheid • Is de classificatie makkelijk te begrijpen en makkelijk en zonder speciale training uitvoerbaar?
Constructvaliditeit • Discrimineert de classificatie groepen van elkaar op een manier die relevant is voor het doel? • Vertonen de afzonderlijke groepen relaties met factoren waarmee zij een relatie zouden moeten vertonen?	Reproduceerbaarheid • Is de classificatie reproduceerbaar in een nieuwe groep patiënten in dezelfde setting? • Zijn de variabelen (symptomatologie) waarop de classificatie is gebaseerd reproduceerbaar vast te stellen?
Contentvaliditeit • Is het domein goed gespecificeerd en zijn de relevante patiëntgroepen	Generaliseerbaarheid • Is de classificatie reproduceerbaar in een andere setting

over en weer uitsluiten. Dit is bijvoorbeeld een probleem bij de subtitels binnen de rubriek 'L15 – kniesymptomen/klachten' van de ICPC.^{12,13}

- *Face-validiteit* betreft de klinische herkenbaarheid van de classificatie. De symptomatologie van de clusters binnen de classificatie van heupklachten werd voorgelegd aan twintig experts; zeven van de negen clusters werden herkend en door hen benoemd als klinische syndromen. Ook vertoonden deze clusters vergelijkbare symptomatologie als in de literatuur beschreven syndromen.

- *Stovner*¹⁴ en *Felson & Anderson*⁹ noemen tevens de *predictieve validiteit* als belangrijke factor. Voor de klinische toepasbaarheid van een classificatie in de huisartspraktijk is dit een van de belangrijkste aspecten. Veel classificaties binnen de NHG-standaarden zijn vanuit dit doel ontworpen. Predictieve validiteit omschrijft in welke mate de classificatie het mogelijk maakt de prognose en/of het effect van therapie voor de onderscheiden groepen te voorspellen. Omdat het follow-up-onderzoek nog niet is afgerond, kan over de predictieve waarde van de classificatie van heupklachten nog geen uitspraak worden gedaan.

- Daarnaast moet een classificatie natuurlijk bruikbaar en reproduceerbaar zijn, én moet duidelijk zijn wat het doel van de classificatie is.¹¹

Voor de *bruikbaarheid* in de klinische praktijk van de classificatie van heupklachten zal men moeten nagaan of de classificatiecriteria tijdens een klinisch consult gemakkelijk kunnen worden vastgesteld.

De *reproduceerbaarheid* van een classificatie zal moeten worden getest door na te gaan of de classificatie reproduceerbaar is in nieuwe, vergelijkbare onderzoekspopulaties. Bovendien zal de intrawaarner- en interwaarnerreproduceerbaarheid moeten worden vastgesteld; deze is vooral afhankelijk van de reproduceerbaarheid van de door de arts vast te stellen symptomatologie.

Voor de *generaliseerbaarheid* zal men moeten nagaan of de classificatie reproduceerbaar is in andere settings.

Voordelen en gevaren

Een groot deel van de huidige classificaties heeft als doel te komen tot gestandaardiseerde categorieën van aandoeningen die gebruikt kunnen worden voor het vaststellen van incidentie en prevalentie. De ICPC-codes zijn hiervan een voorbeeld. Voor veel van de categorieën binnen deze classificatie bestaan nog geen gestandaardiseerde beschrijvingen of classificatiecriteria. Bij het indelen van patiënten in dergelijke categorieën blijkt dan ook een hoge variatie tussen huisartsen te bestaan.¹⁵⁻¹⁷ Wanneer standaardisatie en validering ontbreken, streeft men het eigenlijke doel voorbij en creëert men een schijnzekerheid over het voorkomen van de aandoeningen. Nu in de ICPC-2 uitgebreidere classificatiecriteria worden aangedragen, zal de reproduceerbaarheid wellicht verbeteren.¹⁸ De validiteit van deze classificatiecriteria is echter onbekend en een juiste schatting van het voorkomen van aandoeningen is daardoor nog steeds niet gewaarborgd.

Hoewel een indeling naar aandoeningen universeel (setting-overstijgend) kan zijn, geldt dit niet voor classificatiecriteria. Wanneer men classificatiecriteria benoemt zullen deze afhankelijk zijn van het spectrum aan aandoeningen en de onderlinge verdeling daarvan én van de verschijningsvorm van de aandoening binnen de onderzochte populatie. Wanneer classificatiecriteria benoemd en getoetst zijn in een bepaalde populatie, wil dit niet zeggen dat zij in een andere setting ook valide zijn. De ACR-criteria voor artrose van de knie en van de heup zijn bijvoorbeeld gemaakt met patiënten die reumatologische klinieken bezochten, maar zij werden gelanceerd als universele classificatiecriteria. De criteria voor artrose van de heup bleken echter niet valide in de huisartspraktijk¹⁹ en de criteria voor artrose van de knie bleken niet valide in de open populatie.²⁰ Bij elke setting behoren dus specifieke criteria.

De ACR-criteria voor artrose van de heup hadden moeten dienen voor zowel het rapporteren van de incidentie en prevalentie van deze aandoening als voor

'case-definitie' bij interventieonderzoek. Classificatiecriteria kunnen echter verschillen naarmate het doel verschillend is. Theoretisch kan men dan ook verschillende criteria gebruiken voor incidentieonderzoek en voor inclusie in interventieonderzoek. Bij incidentieonderzoek streeft men naar een goede puntschatting van het voorkomen van een aandoening; bij interventieonderzoek wil men misschien liever classificatiecriteria met een hoge specificiteit, zodat er meer zekerheid is dat de ingesloten patiënten de betreffende ziekte hebben en men dus het te verwachten effect bij patiënten met deze aandoening beter kan schatten. Voor het stellen van de diagnose bij een individuele patiënt zal men zeker geen criteria willen gebruiken die slechts een goede puntschatting geven van het voorkomen van een aandoening. In dat geval moet rekening worden gehouden met de consequenties van een foute classificatie. Pas wanneer (classificatie-) criteria zijn ontworpen voor het individuele consult, zou men deze criteria diagnostische criteria mogen noemen. In het verlengde hiervan moet men zich ook altijd afvragen met welk doel classificatiecriteria zijn gemaakt wanneer ze worden overgenomen door bijvoorbeeld verzekeringsinstanties.

Voor het klinisch handelen van de huisarts is een classificatie zuiver op grond van prognose, behandeling en verwijzing de meest pragmatische indeling. Vaak zullen de verschillende aandoeningen hierin herkenbaar zijn, omdat bij de diverse aandoeningen prognose en behandeling regelmatig verschillen. Het kan echter ook zijn dat twee verschillende aandoeningen (vooralsnog) dezelfde prognose hebben en/of dezelfde behandeling behoeven; in een pragmatische classificatie zou men deze twee aandoeningen dan niet meer apart onderscheiden. Daarnaast kunnen andere belangrijke, niet-aandoeninggerelateerde factoren (sociaal, mentaal en fysiek) in een dergelijke indeling een belangrijke rol gaan spelen. Vanzelfsprekend zal de huisarts deze kennis willen gebruiken in de dagelijkse praktijk. Toch moet men niet uit het oog verliezen dat een op aandoeningen

gebaseerde indeling onmisbaar is voor het genereren en toepassen van nieuwe kennis, en voor het vaststellen van incidentie en prevalentie. Men kan verschillende classificaties naast elkaar laten bestaan, mits men het doel van de classificatie steeds goed voor ogen houdt.

Literatuur

- 1 King LS. Medical thinking. A historical preface. Princeton / New Jersey: Princeton University Press, 1982.
- 2 Wulff HR, Pedersen SA, Rosenberg R. Philosophy of medicine – an introduction. Oxford: Blackwell Scientific Publication, 1986.
- 3 Bloch DA, Moses LE, Michel BA. Statistical approaches to classification. *Arthritis Rheum* 1990; 33: 1137-44.
- 4 Romesburg HC. Cluster analysis for researchers. Belmont/California: Lifetime Learning Publications, 1984.
- 5 De Jongh AC. Schouderandoeningen in de huisartspraktijk [Dissertatie]. Rotterdam: Erasmus Universiteit Rotterdam, 1994.
- 6 Winters JC, Groenier KH, Sobel JS, et al. Classification of shoulder complaints in general practice by means of cluster analysis. *Arch Phys Med Rehab* 1997;76:1369-74.
- 7 Bierma-Zeinstra SMA. Hip pain in general practice. Exploration and classification [Dissertatie]. Rotterdam: Erasmus Universiteit Rotterdam, 1999.
- 8 Altman RD, Alacron G, Applerouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arth Rheum* 1991;34(5):505-14.
- 9 Felson DT, Andersson JJ. Methodological and statistical approaches to criteria development in rheumatic diseases. *Baill Clin Rheum* 1995;9: 253-66.
- 10 Milligan GW, Cooper MC. An examination of procedures for determining the number of clusters in a data set. *Psychometrica* 1985;50:159-79.
- 11 Buchbinder R, Goel V, Bombardier C, Hogg-Johnson S. Classification systems of soft tissue disorders of the neck and upper limb: do they satisfy methodological guidelines? *J Clin Epidemiol* 1996;49:141-9.
- 12 Lamberts H, Wood M. ICPC International Classification of Primary Care. Oxford: Oxford Medical Publications, 1987.
- 13 Gebel RS. De ICPC-1. Stand van zaken. Praktische problemen en oplossingen. *Huisarts Wet* 2000;43:20-2, 28.
- 14 Stovner LJ. The nosologic status of the whiplash syndrome: A critical review based on a methodological approach. *Spine* 1996;21:2733-46.
- 15 Bierma-Zeinstra SMA, Lipschart SS, Njoo KH, et al. How do general practitioners manage hip problems in adults? *Scand J Prim Health Care* 2000;18:159-64.
- 16 Westert GP, de Bakker DH, Schellevis FG. Interdokter- en interpraktijkvariatie binnen en tussen huisartsregistraties. Een secundaire analyse. *Huisarts Wet* 1999;42:18-21.
- 17 Hofmans EA. Diagnostische interdoktervariatie. Kantekeningen bij een taai ongerief. *Huisarts Wet* 1999;42:3-4.
- 18 ICPC-2. International Classification of Primary Care. Second edition. Oxford, etc.: Oxford University Press, 1998.
- 19 Bierma-Zeinstra SMA, Bohnen AM, Ginai AZ, et al. Validity of ACR criteria for diagnosing hip osteoarthritis in primary care research. *J Rheumatol* 1999;26:1129-33.
- 20 Schouten SAG, Valkenburg HA. Classification criteria: Methodological considerations and results from a 12 year following study in the general population. *J Rheumatol* 1995;22(Suppl 43): 44-5.

H&W lezen is werk... maar wel leuk werk

Literatuur (vervolg van pag. 457)

- 1 Bouter LM, Knottnerus JA. Beroordeling van de maatschappelijke relevantie van toegepast gezondheidsonderzoek: het belang van publiceren in nationale vaktijdschriften als ruwe indicator. *Ned Tijdschr Geneesk* 2000;144:1178-83.
- 2 Medicoll'98, bereiksonderzoek van medische vakbladen bij huisartsen. Hilversum: Groep Vaktijdschriften, Nederlands Uitgeversverbond, Collectief Medische Vakbladen, 1998.
- 3 Meijman FJ. De meningen van lezers over Huisarts en Wetenschap. *Huisarts Wet* 1991;34:252-6.
- 4 Boot CPM, Meijman FJ. Systematische preventie van hart- en vaatziekten in de huisartspraktijk is niet zinvol. *Huisarts Wet* 1999;42:341-6.

Lactosemalabsorptie. Oorzaken en klinische consequenties

Vervolg literatuur pag. 468

- 24 Barr RG, Levine MD, Watkins JB. Recurrent abdominal pain of childhood due to lactose intolerance. *N Engl J Med* 1979;300:1449-52.
- 25 Hyams JS, Treem WR, Justinich CJ, et al. Characterization of symptoms in children with recurrent abdominal pain: resemblance to irritable bowel syndrome. *J Pediatr Gastroenterol Nutr* 1995;20:209-14.
- 26 Suarez FL, Savaiano DA, Levitt MD. A comparison of symptoms after the consumption of milk or lactose-hydrolyzed milk by people with self-reported severe lactose intolerance. *N Engl J Med* 1995;333:1-4.
- 27 Kneepkens CMF, Hoekstra JH. Peuterdiarree. *Ned Tijdschr Geneesk* 1996;140:2026-8.
- 28 Lisker R, Aguilar L, Zavala C. Intestinal lactase deficiency and milk drinking capacity in the adult. *Am J Clin Nutr* 1978;31:1499-503.
- 29 Vesa TH, Korpela RA, Sahi T. Tolerance to small amounts of lactose in lactose maldigesters. *Am J Clin Nutr* 1996;64:197-201.
- 30 Solomons NW, Guerrero AM, Torun B. Dietary manipulation of postprandial colonic lactose fermentation: 1. Effect of solid foods in a meal. *Am J Clin Nutr* 1985;41:199-208.
- 31 Kolars JC, Levitt MD, Aouji M, Savaiano DA. Yogurt – an autodigesting source of lactose. *New Engl J Med* 1984;310:1-3.