

Een experimenteel expertsysteem voor differentiele diagnostiek van het dementiesyndroom

Citation for published version (APA):

Plugge, L. A., Verhey, F. R. J., & Jolles, J. (1989). Een experimenteel expertsysteem voor differentiele diagnostiek van het dementiesyndroom. *Tijdschrift voor Gerontologie en Geriatrie*, 20, 51-58.

Document status and date:

Published: 01/01/1989

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Een experimenteel expertsysteem voor differentiële diagnostiek van het dementiesyndroom*

Een eerste evaluatie

L.A. Plugge, F.R.J. Verhey en J. Jolles**

Samenvatting. EVINCE-I*** is een expertsysteem voor neuropsychiatrische diagnostiek, geïmplementeerd op een personal computer. Het is ontwikkeld om de mogelijkheden te verkennen van een intelligent computerprogramma dat de medisch-besluitkundige procedures en data-analyse van een neuropsychiater simuleert op het gebied van differentiële diagnostiek van het dementiesyndroom. De kennis van het systeem richt zich op de detectie van de ziekte van Alzheimer, multi-infarctdementie en depressie in samenhang met dementie. In een experiment werden EVINCE-I en de menselijke expert vergeleken in de diagnostiek van 29 patiënten, van wie 19 met een vroeg stadium van dementie en 10 met diverse stoornissen, maar zonder dementie. Uit de resultaten blijkt dat EVINCE-I in staat is dementie en de belangrijkste oorzaken daarvan in een vroeg stadium en in overeenstemming met een menselijke expert te diagnostiseren.

Summary: An experimental expert system for differential diagnosis of dementia. EVINCE-I is an expert system for neuropsychiatric diagnostics implemented on a personal computer. It is developed to assess the possibilities of an expert system which models the procedures used for medical decision taking and data analysis by a neuro-psychiatrist when making a differential diagnosis in the domain of demential syndromes. The knowledge of the system centers around detection of Alzheimer's disease, multi-infarct dementia and depression-induced dementia. EVINCE-I and the human expert were compared in their ability to diagnose 29 patients, i.e. 19 patients with early stages of dementia and 10 patients showing varying disorders except dementia. It is shown that EVINCE-I is able to detect dementia and its main causes, and produce diagnoses that agree with those of the human expert.

Trefwoorden: dementiesyndroom, differentiële diagnostiek, expertsysteem, neuropsychiatrie, technology assessment (TA). Tijdschr Gerontol Geriatr 1989; 20: 51-8.

Inleiding

Klachten over een verminderd cognitief functioneren komen bij een groot aantal zeer uiteenlopende ziektebeelden voor, terwijl ook de normale veroudering gepaard gaat met een afname van cognitieve functies.^{1,2} In het bijzonder bij ouderen komen geheugenklachten vaak voor. Deze zogenaamde 'mil-

de ouderdomsvergeetachtigheid' is moeilijk te onderscheiden van die soorten geheugenstoornissen die een voorbode kunnen zijn van een hersenziekte, bijvoorbeeld Alzheimer dementie (AD), of die een uiting zijn van ernstige psychopathologie, bijvoorbeeld depressie.³ Mede daarom is de differentiële en vroegdiagnostiek van deze ziekten een lastige onderneming.⁴⁻⁷ Het probleem wordt vergroot door het feit dat de criteria voor het begrip dementie niet duidelijk zijn. Zo wordt het door sommigen gebruikt in de betekenis van dementiesyndroom, door anderen wordt het gezien als een nosologische (ziekte)eenheid en soms wordt het zelfs als synoniem voor de ziekte van Alzheimer gebruikt.⁸

Met de introductie van DSM-III in 1980 en de criteria van de NINCDS-ADRDA-consensuswerkgroep in 1984 is er enige systematiek gekomen in de classificatie van ziektebeelden die met symptomen van dementie gepaard gaan.^{9,10} De diagnostiek zelf is er echter nog niet mee verbeterd. Een groot probleem hierbij betreft het feit dat de diagnose 'waarschijnlijk ziekte van Alzheimer' grotendeels een diagnose per exclusionem is. Dat wil zeggen dat alle andere mogelijkheden moeten zijn verworpen voordat deze diagnose kan worden gesteld. (De definitieve diagnose 'ziekte van Alzheimer' kan overigens pas op grond van neuropathologisch onderzoek gesteld worden.) Dit alles betekent dat een grote hoeveelheid zeer verschillende deelonderzoeken moet worden uitgevoerd. Het gaat in dezen om CT-scan, EEG, bloedonderzoek, hormoonbepalingen, neuropsychologisch onderzoek naar geheugen-, taal- en andere cognitieve functies, psychiatrisch en neurologisch onderzoek. Ook een uitgebreide anamnese en hetero-anamnese behoren bij het onderzoek. De keuze van de verschillende hulponderzoeken en de weging van de resultaten kunnen slechts door een expert goed worden uitgevoerd. Ook de integratie van de afzonderlijke deelbevindingen is slechts mogelijk op grond van veel klinische ervaring. Deze expertise bestaat uit zowel expliciete als impliciete kennis, die bovendien per expert kan verschillen, zodat het verkrijgen van inzicht in en het overdragen van die kennis tussen experts onderling wordt bemoeilijkt.

Deze problemen brengen met zich mee, dat:

1. een door de vergrijzing van de Nederlandse be-

* Dit onderzoek werd gesubsidieerd door het Ministerie van Welzijn, Volksgezondheid en Cultuur in het kader van de medische Technology Assessment. (TA-87-19) (CRO-236509)

** Resp. cognitief psycholoog, zenuwarts en neuropsycholoog/neurochemicus, verbonden aan de vakgroep Neuropsychologie en Psychobiologie van de Rijksuniversiteit Limburg.

*** EVINCE [i'vins] is afgeleid van het Engelse werkwoord 'to evince' i.e. bewijzen, (aan)tonen. De Romeinse I staat voor het versienummer.

- volking snel groeiende groep patiënten onvoldoende kan worden onderzocht;
2. hulpverleners uit de eerste lijn niet voldoende kunnen inschatten op grond van welke criteria een patiënt moet worden doorverwezen en welke vormen van diagnostiek door de huisarts kunnen worden uitgevoerd;
 3. Het beleid voor het aanvragen van hulponderzoeken niet altijd goed doordacht is en dat zulks leidt tot hogere kosten voor onderzoeken die niet bijdragen aan de diagnostiek en tevens een onnodige belasting vormen voor de patiënt.

Het onderzoeksproject, dat in deze publikatie wordt beschreven, houdt zich bezig met het ontwikkelen van een intelligent computerprogramma dat de hiervoor genoemde expertise bezit. Een dergelijk programma noemt men een expertsysteem (ES) en heeft om de volgende redenen grote waarde voor dit toepassingsgebied:

In de eerste plaats vereist het ontwerpen van een ES dat de expertise wordt gemobiliseerd in de vorm van regels, concepten en criteria waardoor de toegankelijkheid en inzichtelijkheid voor andere experts en niet-experts worden bevorderd.

In de tweede plaats gaat het bij dit type diagnostiek om het verwerken en integreren van een grote hoeveelheid zeer verschillende gegevens uit het diagnostisch en aanvullend onderzoek. Aangezien de computer principieel betere mogelijkheden tot ordenen en sorteren heeft dan de humane (arts-)expert, zou een ES ook in dezen een belangrijke bijdrage kunnen leveren.¹¹

In de derde plaats is een ES op te vatten als een model waarmee simulaties kunnen worden verricht. Gezien de al jaren durende discussie over het belang van de verschillende diagnostische onderzoeken die uitgevoerd dienen te worden bij dementiediagnostiek is dit een belangrijke optie. De bijdrage die de verschillende deelonderzoeken leveren aan de eindiagnose en derhalve aan het op basis daarvan te voeren beleid kan immers simpelweg worden onderzocht door deze variabelen in het ES te manipuleren.

Hoewel er in het laatste decennium veel ES-prototypen zijn ontwikkeld, is de psychiatrische expertise door haar complexiteit zeer zelden onderwerp van ES-ontwikkeling geweest.¹² Om de waarde op het gebied van de differentiële diagnostiek van het dementiesyndroom te kunnen inschatten, is het van groot belang dat aangetoond kan worden dat een dergelijk ES diagnoses produceert die gelijk zijn aan die van de domeinexpert. Dit artikel beoogt, met behulp van het ES EVINCE-I, aan te tonen dat een ES op het gebied van dementie goede resultaten oplevert.

De ontwikkeling van EVINCE-I

Allereerst wordt in het kort besproken wat een ES is, waarna een verantwoording volgt voor de keuze van het in dit project gebruikte systeem. Daarna zal een algemeen beeld worden gegeven van de in EVINCE-I opgeslagen kennis, door de beslisprocedures stap voor stap te volgen.

Expertsystemen: Basisprincipe

Een ES is een computerprogramma dat tot taak heeft de prestaties van een menselijk expert op één bepaald gebied ten minste te evenaren en niet-experts de mogelijkheid biedt de experts te consulteren zonder dat deze aanwezig hoeft te zijn. Daarnaast kan een ES gebruikt worden door een expert om tijd te winnen en/of alle mogelijke uitkomsten van een probleemstelling te verkrijgen.

ES-en kunnen op verschillende manieren functioneren, maar de meest gebruikelijke is een programma dat bestaat uit vier onderdelen:^{13 14}

- een kennisbestand (knowledge-base), bestaande uit een regel- (rule-base) en een conceptenbestand (te vergelijken met een database);
- een redeneermechanisme (inference engine) om met behulp van feiten en regels nieuwe feiten te herleiden, bijv. een diagnose;
- een uitlegmechanisme om het systeem te vragen hoe en waarom het bepaalde conclusies heeft getrokken;
- een dialoogmechanisme (user interface), waarmee de communicatie tussen gebruiker en systeem wordt verzorgd.

Naast de ES-en, die zich beperken tot één kennisdomein, zijn er zogenaamde ES-shells ontwikkeld. Shells onderscheiden zich van ES-en door het ontbreken van de domeinspecifieke kennis, de knowledge-base, en vormen het geraamte van ES-en. Op deze manier kan één ES relatief eenvoudig aangepast worden aan de eisen van verschillende probleemgebieden door het bouwen van een nieuwe knowledge-base.

De keuze van de ES-shell

Een van de belangrijkste beslissingen in het ontwikkelen van een ES is de keuze van de juiste ES-shell.¹³

¹⁷ Het type shell bepaalt welke kennisrepresentatie en inferenties mogelijk zijn, omdat deze technieken binnen een shell vastliggen. Verder stelt het domein van expertise een aantal duidelijke eisen. In dit geval betekende het dat een lerend systeem niet wenselijk zou zijn. Een dergelijk systeem probeert zelfstandig regels te induceren met behulp van gekregen voor-

beelden, waardoor menselijke controle over de juistheid van de kennis minimaal zou zijn. Bovendien zijn lerende systemen sterk afhankelijk van het soort voorbeelden dat gegeven wordt en de volgorde waarin ze gegeven worden. Ten tweede betekende het dat shells met een puur statistische aanpak onwenselijk zouden zijn omdat zo'n aanpak geen direct interpreteerbare regels oplevert.¹⁸ Een systeem gebaseerd op produktieregels in de vorm van *als: X en Y dan: Z* en een inferentiemechanisme dat zowel data – als hypothese – georiënteerd zou kunnen werken, zou het meest geschikt zijn. Dit laatste criterium sloot shells uit die maar één type van redeneren kennen. Van de overgebleven shells is uiteindelijk gekozen voor ACQUAINT, omdat het kennisbestand gepresenteerd wordt in de vorm van frames, vergelijkbaar met de systematiek van een database.^{15,16} Verder wordt procedurele kennis gerepresenteerd met behulp van regels en metaregels in een *als-dan* vorm en is het inferentiemechanisme in staat tot zowel voorwaarts als achterwaarts redeneren, d.w.z. zowel data- als hypothesegestuurd. Bovendien maakt het inferentiemechanisme naast de in de logica gebruikelijke uitspraken als 'waar' en 'onwaar', gebruik van onzekerheid in de vorm van zekerheidsfactoren (CF, Certainty Factor).*

ACQUAINT draait op een standaard IBM-compatibele machine met minimaal 512 K-RAM en twee floppy drives. Een andere aantrekkelijke eigenschap is, dat de eindgebruiker een kant en klaar programma krijgt (een 'run-time copy') en niet het hele ontwikkelsysteem ACQUAINT.

Opbouw van het kennisbestand

Voor het ontwikkelen van het kennisbestand werd gebruik gemaakt van de internationaal geaccepteerde criteria zoals beschreven in het DSM-III en het rapport van de NINCDS-ADRDA-consensuswerkgroep.^{8,9,10} Een zenuwarts, verbonden aan de vakgroep Neuropsychologie en Psychobiologie van de Rijksuniversiteit Limburg, fungeerde als de domeinexpert die de expertise in het gebruik van bovenstaande criteria en onderzoeksprocedures verwoordde. (Zie voor een nadere uiteenzetting referenties 8 en 19.)

Na een voorbereidende fase om een algemeen beeld te verkrijgen van het kennisdomein, werd de probleemruimte verder afgebakend. Daarvoor wer-

den bij de domeinexpert interviews afgenomen over de procedures en gegevens die gebruikt worden in een onderzoek. Uit de resulterende protocollen werd de algemene onderzoeksprocedure van de domeinexpert opgesteld en opgesplitst in onderdelen, gebaseerd op de eerder genoemde deelonderzoeken (zie pag. 51). Deze onderdelen werden vervolgens onderzocht op hun onderlinge samenhang en de daarin gebruikte begrippen (concepten), regels en criteria. Aan de hand van de protocollen werd de informatie ingedeeld in drie typen relevantie:

1. essentiële informatie,
2. informatie die belangrijk is naar aanleiding van andere informatie,

Tabel 1a. Patiëntgegevens gebruikt door EVINCE-I. (Patient characteristics used by EVINCE-I).

Onderwerp	Inhoud
Persoonsgegevens:	leeftijd, patiëntnaam, patiëntnummer, geboortedatum, consultdatum, sekse, handvoorkeur, beroep, sociaal-economische achtergrond, opleiding.
Haematologie:	BSE, Hb, Ht-percentages, Leuco's.
Chemie:	Kalium, Na, Ca, P, Glucose, B ₁ , B ₁₂ , AF, Gamma-GT, OT, PT, T ₄ , TSH.
Aanvullend onderzoek:	EEG, CT-scan, ECG.
Anamnese:	gewicht, alcoholgebruik (nu en vroeger), eetlust, tabakgebruik (nu en vroeger), intoxicatie, beroepsfunctioneren, sociaal-functioneren.
Psychologische en cognitieve functies:	geheugen, abstracte denken, hogere corticale functies (fatische, praktische en gnostische functies), bewustzijn, aandacht, oordeel, oriëntatie, perceptie.
Motoriek:	spraak, motorische activiteit.
Persoonlijkheid:	karakter, persoonlijkheid, sociale activiteit, ziekte-inzicht.
Klachtpatroon:	begin, wijze van achteruitgang, beloop, lichamelijke klachten.
Slaap:	slaperigheid, uitslapstoornis, doorslapstoornis, uitslapstoornis, onrust.
Stemming:	labiliteit, fluctuatie, dagschommelingen, schuldgevoelens, suicide, remmingen, agitatie, angst met psychologische of lichamelijke verschijnselen (algemeen, gastro-intestinaal), libido, hypochondrie, depersonalisatie, depersonalisatietype, paranoïde symptomen, dwangsymptomen, Hamiltondepressiescore.
Atherosclerose:	anamnestische hypertensie, TIA, CVA, atherosclerose, focaal neurologische symptomen, focaal neurologische tekenen, Hachinskiscore.

In bovenstaande lijst zijn de variabelen waarmee EVINCE-I wordt gecontroleerd en gestuurd op programmanivo niet opgenomen, omdat ze slechts indirect met de patiëntgegevens te maken hebben en voor de gebruiker meestal verborgen blijven.

* Voor een uiteenzetting betreffende de berekening van zekerheidswaarden (CF's) in ACQUAINT, wordt verwezen naar referentie 15 (pp. 1-21 e.v.). Een meer algemene uiteenzetting met voorbeelden over CF's, is te vinden in referentie 14 en 17. Het gebruik van onzekerheid (fuzzy-logica) in ACQUAINT wordt uiteengezet in referentie 15 (pp. 5-23).

Tabel 1b. Diagnostische kennis van EVINCE-I. (Diagnostic knowledge used by EVINCE-I).

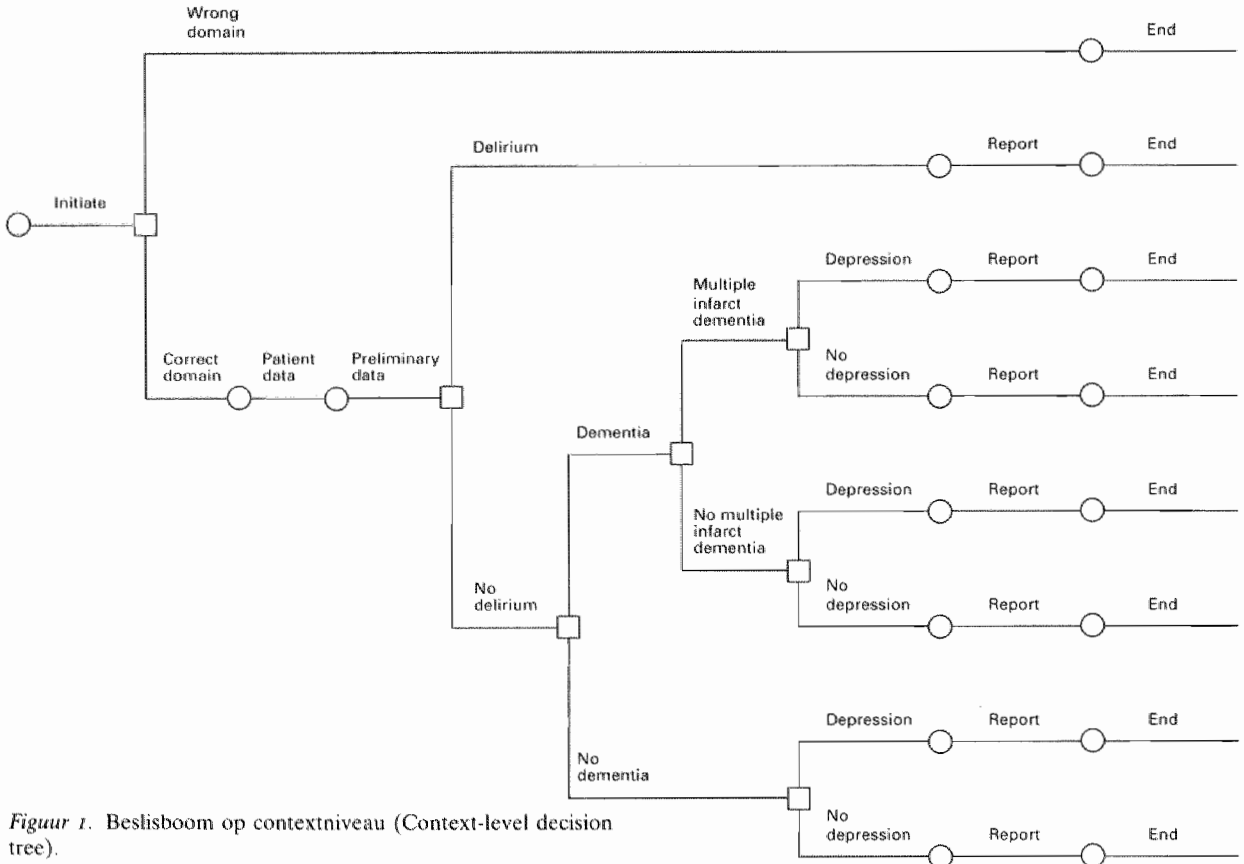
- Dementiesyndroom (seniel/pre-seniel)
- Primair degeneratieve dementie (seniel/pre-seniel)
- Multiple infarct-dementie
- Depressie in engere zin
- Delier
- Cognitieve Stoornissen
- Schilddklierstoornis (verhoogde/verlaagde werking)
- Diabetes
- Alcoholisme
- Electrolietstoornissen
- Ontsteking
- Vitamine-B-insufficiënties
- Medicatie-invloed
- Epilepsie

3. informatie waarvan de relevantie nog niet duidelijk is.

Van ieder concept werd de domeinexpert gevraagd aan te geven welke onderzoeksvraag centraal staat en welke informatie het concept oplevert. Zie voor een opsomming van de patiëntgegevens tabel 1a en voor de diagnostische kennis tabel 1b.

De meerderheid van die concepten is terug te vinden in internationaal geaccepteerde criteria zoals de Hamilton Depression Rating Scale (HDRS), de Hachinski-index en het DSM-III.^{20 21 9} Deze gegevens konden zonder veel problemen gegroepeerd worden

in contexten, i.e. groepen regels en gegevens die een onderlinge samenhang hebben, analoog aan die waar ze uit voortkwamen. Het overgrote deel van deze gegevens werd verzameld door de domeinexpert. De overige gegevens werden verzameld door andere specialisten (bloed- en urineonderzoek, ECG, EEG, CT-scan en de thorax-röntgenfoto). Deze doen dienst als aanvullend onderzoek, waarvan een groot deel routinematig wordt gedaan. De gegevens daarvan waren moeilijk in het besluitvormingsproces in te delen, omdat ze niet bij één specifiek onderzoeksgebied, c.q. context, horen. Routinematig bloedonderzoek wordt bijvoorbeeld uitgevoerd om er zeker van te zijn dat een aantal oorzaken van het dementiesyndroom kan worden uitgesloten, bijvoorbeeld een vitamine B₁₂-deficiëntie. EEG en CT-scan worden met eenzelfde doel afgenomen. Ook hiermee kunnen bepaalde oorzaken uitgesloten worden en is het mogelijk bepaalde afwijkingen te constateren zoals een intracerebrale tumor, epilepsie of infarct. Bovendien helpen deze gegevens bij het bepalen van de richting voor verder onderzoek en kunnen ze andere bevindingen bevestigen. Dit laatste geldt overigens ook voor sommige gegevens die door de domeinexpert verzameld worden, zoals onderdelen van de anamnese, heteroanamnese en het neurologisch onderzoek.



Figuur 1. Beslisboom op contextniveau (Context-level decision tree).

Aangezien een zoekstrategie werd gewenst die analoog is aan die van de domeinexpert, was het belangrijk dat het ES de gegevens op een analoge manier verwerkt. Omdat de routine-gegevens als eerste worden verzameld, zijn deze concepten ondergebracht in de context 'Preliminary Data', terwijl de overige bij hun specifieke deelonderzoeken zijn ondergebracht.

De beslisprocedures

De 'Initiate Context' is de hoofdcontext die alle andere omvat en een introductie in het systeem geeft. In de volgende context beslissen een paar regels of het onderzoeksprobleem valt onder de beschikbare kennis. Als dat het geval is, of als het onzeker is, start het systeem met de eigenlijke consultatie.

De volgende contexten simuleren zo precies mogelijk de door de domeinexpert gehanteerde procedure bestaande uit doelen, en stappen om die doelen te bereiken. De eerste stap van het ES is om anamnestic data te verzamelen, zoals geboortedatum, geslacht, het berekenen van de leeftijd, etc.

Het eerste doel daarna is vast te stellen of het een geval van dementie betreft, zonder uitspraak over de oorzaak, omdat 'Dementia is a diagnosis based on behaviour and cannot be determined by CT-scan, EEG or other laboratory instruments, (...)'.¹⁰ In de eerste stap daartoe worden de preliminaire data gecontroleerd op mogelijke afwijkingen die al direct tot een diagnose kunnen leiden. De bevindingen daarvan worden in het verdere onderzoek gebruikt.

In de tweede stap wordt nagegaan of de patiënt een helder bewustzijn heeft. Als dat niet het geval is, d.w.z. hij vertoont symptomen die voldoen aan de DSM-III-criteria voor delier, dan wordt de consultatie beëindigd. Het systeem maakt dan een eindrapport op met het advies de consultatie opnieuw te starten nadat het delier is behandeld. Tevens probeert het systeem de oorzaak voor het delier aan te geven, op grond van de beschikbare gegevens. Het besluit om van verder onderzoek af te zien wordt genomen op basis van het criterium dat er sprake moet zijn van een helder bewustzijn.⁹

Na het constateren van een helder bewustzijn worden in de volgende stap de DSM-III-criteria voor dementie toegepast. Indien er geen dementie wordt gediagnostiseerd zal EVINCE-I proberen na te gaan of een vorm van depressie oorzaak kan zijn van de gerapporteerde klachten. In het eindrapport worden dan alle diagnoses verzameld en gecontroleerd op wederzijdse implicaties.

Voldoen de symptomen echter aan de criteria van dementie, dan is het volgende doel na te gaan wat de oorzaak daarvan kan zijn. Gezien de aard van de

oorzaak met de grootste incidentie, i.e. AD, wordt dit op een speciale manier gedaan. AD wordt voornamelijk gediagnostiseerd m.b.v. exclusiecriteria en het stellen van die diagnose wordt daarom uitgesteld tot het laatst. Als eerste stap om AD uit te sluiten wordt er gekeken naar mogelijke vasculaire problemen die kunnen leiden tot de diagnose multiple infarct-dementie (MID), de tweede hoofdoorzaak van dementie. Onafhankelijk van die uitkomst wordt in de tweede stap gecontroleerd op de aanwezigheid van een depressie. In stap drie wordt er gekeken of er al diagnoses gesteld zijn op basis van de preliminaire gegevens, die de diagnose AD kunnen uitsluiten. Als al deze stappen zijn doorlopen en de diagnose AD nog steeds niet is uitgesloten, wordt m.b.v. de preliminaire gegevens de CF van de diagnose AD aangepast. De CF wordt verhoogd als het EEG-patroon bijvoorbeeld een toegenomen vertraging van het alfaritme laat zien en wordt verlaagd als het EEG tekenen van focale pathologie vertoont. Geen van die gegevens kan de diagnose AD echter uitsluiten, ze worden gebruikt zoals voorgesteld door de NINCDS-ADRDA-consensuswerkgroep: ter verhoging van de waarschijnlijkheid.¹⁰ Overigens maakt EVINCE-I nog geen onderscheid tussen 'possible', 'probable' en 'definite' AD.¹⁰ In een latere versie wordt deze indeling wel geïmplementeerd.

Het eindrapport wordt opgemaakt in de Report-context. Deze simuleert de evaluerende activiteit van de domeinexpert waarin bevindingen worden herformuleerd in standaardzinnen. Een toekomstige versie zal complete standaardrapporten schrijven. Een voorbeeld van een rapport geeft tabel 2.

Tabel 2. Diagnostisch rapport van EVINCE-I. (Diagnostic Report by EVINCE-I).

EVINCE-I	
Report d.d. 27-03-87	
Patient data concerning : X,Y	
Date of birth	: 11-06-10
Age	: 76
Sex	: male
Hand preference	: Right
Occupation	: none
Soc. Econ. Status*	: 4
Educational level**	: 2
Due to the fact that the Hachinski score is 3, there are no clues (i.e. 65%) for a causal connection with vascular problems.	
The Hamilton depression score is 11 points, indicating (with 60% certainty) a mild form of depression.	
Patient X suffers, with a certainty of 89%, from dementia with a senile onset.	
Primary Degenerative Dementia is likely with a certainty of 65%.	

(De initialen en geboortedatum van patiënt zijn veranderd.)

* Sociaal Economische Status, indeling naar Diesfeldt.²¹

** Opleidingsniveau, systeem Verhage.²²

Een vergelijking tussen EVINCE-I en de domeinexpert

Introductie

Zoals reeds opgemerkt was het doel na te gaan of een desk-top ES krachtig genoeg is om valide diagnoses te stellen. De prestaties van EVINCE-I werden daarom vergeleken met die van de domeinexpert. Met behulp van het experiment werden de volgende hypothesen getoetst:

1. de domeinexpert en EVINCE-I komen overeen in hun diagnoses,
2. de domeinexpert en EVINCE-I komen overeen in de zekerheid van hun diagnoses.

Verder werd de soort fout-positieve en fout-negatieve diagnoses van EVINCE-I met die van de expert vergeleken, om te zien of het ES bepaalde typen fouten maakte. Indien de hypothesen bevestigd zouden worden, zou dat betekenen dat de juiste kennis aan de domeinexpert is ontleend en geïmplementeerd en dat die kennis door EVINCE-I op de juiste manier gebruikt wordt.

Methode

Uit het beschikbare patiëntenbestand van de Maas-trichtse Geheugenkliniek werden 19 patiënten gekozen die volgens de inschatting van de domeinexpert met tenminste 50% zekerheid een vorm van dementie vertoonden.¹⁹ Dit om na te gaan of EVINCE-I fout-negatieve dementiediagnosen zou maken. Om de mogelijkheid van fout-positieve diagnoses na te gaan werden bovendien aselekt 10 andere patiënten uit het bestand gekozen die niet de diagnose dementie hadden. De meeste van hen leden aan een milde depressie. De gemiddelde leeftijd van de eerste groep bedroeg 73 jaar (sd = 7.5) en van de tweede 47 (sd = 14.6). Geen van de toegepaste diagnostische criteria maakt gebruik van de leeftijd als onderscheidende factor, ze wordt hier slechts voor de volledigheid vermeld. Alle patiënten waren gediagnostiseerd door de domeinexpert en geen van de gevalsbeschrijvingen is gebruikt voor testen gedurende het ontwikkelen van het systeem.

De expert ontving een formulier met een korte instructie waarin hem gevraagd werd de diagnoses van iedere patiënt nogmaals in sleutelwoorden op te schrijven, met daarnaast een zekerheidspercentage. Deze waarde kon variëren van 0 (onbekend) tot 100 (absoluut zeker), omdat EVINCE-I ook waarden gaf in die range. De data van de 29 patiënten werden door de knowledge-engineer ingevoerd. Alle data die EVINCE-I nodig had waren aanwezig in de patiëntendossiers.

Resultaten

De belangrijkste diagnoses waren: dementie (DEM), dementie van het Alzheimerstype (DAT), multiple infarctdementie (MID), depressie (DEP) en een restcategorie (RES). Aangezien deze laatste een te grote variatie aan diagnoses bevatte voor een zinvolle vergelijking, werd zij uit de statistische analyse weggelaten. Elke patiënt kon één of meer van de voorgaande diagnoses krijgen. De categorieën van de door de domeinexpert en EVINCE-I gestelde diagnoses zijn te vinden in tabel 3.

Tabel 3. Diagnostische categorieën. (Diagnostic categories).

cat.	diagnose			
	DEM	DAT	MID	DEP
1	x	x	-	-
2	x	x	-	x
3	x	-	x	-
4	x	-	x	x
5	x	-	-	-
6	x	-	-	x
7	-	-	-	-
8	-	-	-	x

Legenda: DEM = Dementie
 DAT = Dementie van het Alzheimer type
 MID = Multiple Infarct Dementie
 DEP = Depressie
 x = aanwezig
 - = ontbreekt

Tabel 4 laat de verdeling zien van de diagnoses die de domeinexpert en EVINCE-I maakten, na transformatie van de zekerheidwaarden (CF) naar binaire waarden door de CF door 100 te delen en het resultaat normaal af te ronden naar 1 of 0.

In de eerste toets werd (met behulp van de kappa-coëfficiënt) gekeken in welke mate EVINCE-I en de domeinexpert overeenkwamen in alle diagnostische

Tabel 4. Frequentiematrix van diagnostische categorieën. (Frequency matrix of diagnostic categories).

cat.	EVINCE-I:								ALL
	1	2	3	4	5	6	7	8	
1	7	.	1	8
D E	2	.	1	1
O X	3	.	.	2	2	.	.	.	4
M P	4	.	.	1	2	.	.	.	3
A E	5	.	.	1	.	1	.	.	2
I R	6	.	1	1
N T	7	1	1
8	3	6	9
ALL	7	2	5	4	1	0	4	6	29

Ten behoeve van de leesbaarheid zijn de nullen in de matrix vervangen door punten.

Zie voor de samenstelling van de categorieën tabel 3.

Tabel 5. Interbeoordelaar-overeenstemming per diagnostische categorie. (Interrater agreement per diagnostic category).

Toets	Categorieën	Obs.	Exp.	Kapp κ	Z-score
1	1, 2, 3, 4, 5, 6, 7, 8	0.69	0.18	0.62	7.19
2	(1+2),(3+4),(5+6),(7+8)	0.90	0.29	0.85	7.13
3	(1+2+3+4+5+6), (7+8)	1.00	0.55	1.00	4.89
4	(1+2), (3+4+5+6), (7+8)	0.93	0.33	0.90	6.82
5	(1+2+5+6), (3+4), (7+8)	0.93	0.34	0.90	6.78

Alle Z-scores zijn significant: $p < 0.0001$

Zie voor de samenstelling van de categorieën tabel 3.

combinaties.²⁴ Het resultaat daarvan is te vinden in tabel 5, toets 1.

Zoals al uit de ruwe data kan worden opgemaakt, was de mate van overeenstemming tussen de twee beoordelaars niet volledig. De reden daarvoor is te vinden in de categorieën (7+8) en (3+4), die geacht worden een onderscheid te maken tussen depressie en geen-depressie, maar dat slechts in beperkte mate doen. De mate van overeenstemming steeg beduidend toen de DEP-categorie buiten beschouwing werd gelaten, zoals te zien is in toets 2 van tabel 5.

In toets 3 analyseerden we de mate van overeenstemming tussen de twee beoordelaars op de diagnose DEM, dat wil zeggen de categorieën (1+2+3+4+5+6) en (7+8). Dit is een belangrijke diagnose, omdat de diagnose van DAT en MID daarvan afhankelijk zijn. Toets 3 in tabel 5 laat zien dat beide beoordelaars perfect overeenstemden bij deze diagnose. Daar MID en DAT in dit systeem elkaar uitsluitende diagnoses zijn, werd de overeenstemming voor de diagnose DEM met DAT, d.w.z. de categorieën (1+2), (3+4+5+6) en (7+8), en DEM en MID, d.w.z. de categorieën (1+2+5+6), (3+4) en (7+8), afzonderlijk getoetst. In beide gevallen lieten de beoordelaars een hoge mate van overeenstemming zien. (Zie de toetsen 4 en 5 in tabel 5). De resultaten laten zien dat de overeenstemming hoog tot perfect was, wanneer werd gecorrigeerd voor de diagnose DEP. DEP is de enige diagnose die niet volkomen juist door EVINCE-I werd gehanteerd.

Na deze analyse werden de scores door afronding getransformeerd naar een 11-puntsschaal (0-10) om de richting en relatieve grootte van de zekerheid te vergelijken waarmee de diagnoses gesteld waren. Daarvoor werden dezelfde diagnoses gebruikt als hiervoor, maar nu werden ze afzonderlijk getoetst. Voor iedere zekerheidsscore werd de Spearman-Rang-Correlatiecoëfficiënt berekend. (zie tabel 6) In tabel 6 is te zien dat de beoordelaars in hoge mate correleerden in hun zekerheid over de diagnoses, met uitzondering van de diagnose DEP, die een matige correlatie laat zien.

Hoewel de RES-categorie uit de analyse is wegge-

Tabel 6. Interbeoordelaar-overeenstemming met een 11-puntschaal. (Interrater agreement using a 11-points certainty scale).

Category	Spearman's corr:	DF:	t	Prob:
DEM	0.916	27	11.86	$p < 0.001$
DAT	0.821	27	7.47	$p < 0.001$
MID	0.849	27	8.35	$p < 0.001$
DEP	0.560	27	3.51	$p < 0.002$
N = 29				tweezijdig

laten, kunnen we daarin toch enige aanwijzingen vinden over de prestaties van EVINCE-I in het diagnostiseren van meer zeldzame oorzaken van demantie. Eén patiënt kreeg van EVINCE-I de diagnose MID, maar leed in werkelijkheid aan Multiple Sclerose (MS). Aangezien veel symptomen daarvan ook bij MID voorkomen en EVINCE-I het concept MS niet kende, is MID een logisch (fout) alternatief. Bij de diagnose van een andere patiënt gaven zowel EVINCE-I als de domeinexpert het commentaar 'cave vitamine B₁₂-deficiëntie'. In twee andere gevallen werd door beide alcohol als mogelijke oorzaak van het dementiesyndroom opgemerkt. Eén van de twee gevallen waarin de domeinexpert een afwijking in de schildklier opmerkte, werd ook door EVINCE-I opgemerkt.

Discussie

Het feit dat de resultaten voor de categorieën DEM, DAT, MID en de meer zeldzame oorzaken van demantie zo goed zijn, toont aan dat de kennis van EVINCE-I een goede afspiegeling is van menselijke expertise. Wanneer we daarbij in acht nemen dat EVINCE-I daarvoor niet meer dan 122 verschillende concepten (onderzoeksgegevens) nodig heeft, wat substantieel minder is dan de domeinexpert in beschouwing neemt, wordt de prestatie nog interessanter. Het betekent namelijk dat EVINCE-I gebruikt kan worden als middel om de waarde van onderzoeksgegevens te wegen en voor het opstellen van een standaard-onderzoeksprotocol. Het feit dat de overeenkomst op het gebied van depressiediagnostiek minder hoog is dan bij de andere categorieën kan twee oorzaken hebben:

1. Het is mogelijk dat EVINCE-I teveel waarde hecht aan de score van de HDRS. In een evaluatiegesprek met de domeinexpert bleek dat deze de HDRS-score slechts als hulpmiddel gebruikt. Deze mogelijkheid wordt op dit moment onderzocht.
2. Een andere oorzaak zou kunnen liggen in een meta-regel, volgens welke op basis van de stemmingstoestand van de patiënt al dan niet wordt besloten onderzoek te doen naar een mogelijke

depressie en de HDRS-score wordt bepaald, terwijl de domeinexpert dat standaard doet.

De eerste mogelijkheid duidt erop dat EVINCE-I op dit gebied meer informatie nodig heeft, terwijl de tweede erop wijst dat het ES te snel besluit geen onderzoek te doen naar een mogelijke depressie, met als gevolg 4 fout-negatieve depressiediagnosen.

Aangezien de juistheid van een diagnose in dit domein maar bij uitzondering met absolute zekerheid is te stellen en de domeinexpert hier bovendien als gouden standaard fungeerde, is het niet mogelijk met de gegevens van dit experiment na te gaan wie van de twee beoordelaars, de domeinexpert of EVINCE-I, uiteindelijk de juiste diagnosen gesteld hebben.

Slotopmerkingen

Uit de voorgaande discussie bleek dat de kennis en prestaties van EVINCE-I alle aanleiding geven tot verder onderzoek, waarbij het ES kan dienen als gereedschap. Zelfs in die gevallen waarin de diagnose niet optimaal is, kan het ES informatie verschaffen over het onderzochte kennisdomein, omdat nauwkeurig is vastgesteld welke gegevens zijn gebruikt en welke procedure werd gevolgd. Daarnaast geven de resultaten aanleiding het ES te beschouwen als een serieus hulpmiddel voor de expert en andere betrokken medici, voor (vroeg en differentiële) diagnostiek van het dementiesyndroom. Naar aanleiding van de in dit artikel gerapporteerde resultaten is EVINCE-I op een aantal punten bijgesteld, o.a. op het gebied van depressie-onderzoek. Met deze verbeterde versie is een uitgebreid vervolgonderzoek uitgevoerd waarin EVINCE vergeleken wordt met meerdere dementie-experts. De resultaten worden op dit moment geanalyseerd. In een ander onderzoek zullen de bruikbaarheid en betrouwbaarheid getest worden in praktijksituaties, zoals bijvoorbeeld bij huisartsen en psychiaters. Afhankelijk van die resultaten kan begonnen worden met de opzet van een versie voor algemener gebruik in de gezondheidszorg.

Literatuur

1. Jolles J. Cognitive, emotional and behavioral dysfunctions in aging and dementia. In: Swaab DF, Fliers E, Mirmiran M et al (eds.) *Progress in Brain Research*, Amsterdam: Elsevier Science Publishers BV 1986; 70: 15-39.
2. Haase GR. Diseases Presenting as Dementia. In: Wells ChE (ed). *Dementia*. Philadelphia: Davis Comp, 1971, 2nd ed., 27-67.
3. McAllister ThW. Overview: Pseudodementia. *Am J Psychiat*, 1983; 140: 528-33.
4. Jolles J. Early diagnosis of dementia: possible contribu-

tions from neuropsychology. In: Traber J, Gispens WH (eds). *Senile Dementia of the Alzheimer Type*. Berlin: Springer Verlag, 1985; 84-100.

5. Marsden CD. Assessment of dementia. In: Vinken PJ, Bruyn GW, Klawans HL (eds). *Handbook of Clinical Neurology*. Vol. 2. Neurobehavioural Disorders. Amsterdam: Elsevier Sc Publ BV, 1985, 46: 221-232.
6. Marsden CD. Neurological Causes of Dementia other than Alzheimer's Disease. In: Kay DWK, Burrows GD (eds.). *Handbook of Studies on Psychiatry and Old Age*. Amsterdam: Elsevier Sc Publ BV, 1984, 145-67.
7. Van Crevel H. Clinical Approach to Dementia. In: Swaab DF, Fliers E et al (eds.). *Progress in Brain Research*. Volume 70. Amsterdam: Elsevier Sc Publ BV, 1986, 3-13.
8. Verhey FRJ, Jolles J. Over de spraakverwarring rond het begrip dementie en de ziekte van Alzheimer. *Tijdschr Gerontol en Geriatr*, 1988; 19: 89-96.
9. *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition, Washington DC: American Psychiatric Association, 1983.
10. McKhann G, Drachman D, Folstein M, et al. Clinical Diagnosis of Alzheimer's Disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 1984; 34: 939-44.
11. Maxmen JS. Long-term trends in health care: The post-physician era reconsidered. In: Schwefel D (ed). *Indicators and trends in health and health care*. Berlin: Springer Verlag, 1987.
12. Waterman DA. A Guide To Expert Systems, Reading: Addison-Wesley PC, 1986; 272-88.
13. Hayes-Roth F, Waterman DA, Lenat D. *Building expert systems*. Reading: Addison-Wesley PC, 1983, 210-12.
14. Winston PH. *Artificial Intelligence*. Reading: Addison-Wesley PC, 1984, 2nd ed.
15. *ACQUAINT: User's manual for Acquaint and Acquaint-Light*. Manual, Purmerend. Lithp Systems BV, 1987.
16. Tello ER. ACQUAINT. In: Lemmons Ph (ed). *BYTE*. Peterborough: McGraw-Hill Inc, 1987; 7: 265-72.
17. Buchanan BG, Shortliffe EH (eds). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading: Addison-Wesley PC, 1984.
18. Williams T. Computer Aids to Clinical Decision. In: Williams BT (ed). *Computer Aids to Clinical Decision*. Volume 1. Florida: CRC Press Inc, 1982: 63-94.
19. Hamilton M. Development of a Rating Scale for Primary Depressive Illness. *Brit J Soc Clin Psychol*, 1967; 6: 278-96.
20. Hachinski VC, Illif ZE, Zilka E et al. Cerebral bloodflow in dementia. *Arch of Neurol*, 1975; 32: 632-37.
21. Diesfeldt HFA. *De draad kwijt*. Dissertatie, Utrecht, 1983.
22. Verhage F. *Intelligentie en leeftijd*. Dissertatie, Assen: 1964.
23. Verhey FRJ, Reyersens van Buuren EJ, Jolles J. De Geheugenkliniek: Multidisciplinaire benadering bij stoornissen van het geheugen en andere cognitieve functies. In: Houweling ThJ (ed). *Ouder worden nu*. Amsterdam: Gerontologisch Instituut Amsterdam, 1987; 82-8.
24. Schouten JA. *Statistical measurement of interobserver agreement*. Dissertatie, Utrecht: 1985.

Correspondentie: prof. dr. J. Jolles, vakgroep Neuropsychologie & Psychobiologie, RL, Postbus 616, 6200 MD Maastricht, tel. 043-888430.

Ontvangen: september 1988; geaccepteerd: november 1988.