

A desktop expert system for the differential diagnosis of dementia

Citation for published version (APA):

Plugge, L. A., Verhey, F. R. J., & Jolles, J. (1990). A desktop expert system for the differential diagnosis of dementia: an evaluation study. *International Journal of Technology Assessment in Health Care*, 6(1), 147-156. <https://doi.org/10.1017/S0266462300009004>

Document status and date:

Published: 01/01/1990

DOI:

[10.1017/S0266462300009004](https://doi.org/10.1017/S0266462300009004)

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.

A DESKTOP EXPERT SYSTEM FOR THE DIFFERENTIAL DIAGNOSIS OF DEMENTIA

An Evaluation Study

Leonard A. Plugge

Frans R. J. Verhey

Jellemer Jolles

State University of Limburg

Abstract

EVINCE-I is a desktop expert system for the differential diagnosis of dementia, implemented on a personal computer. It is intended to assess the effectiveness of this new technology in modeling a psychiatrist who uses international guidelines for diagnosing dementia. EVINCE-I was tested in diagnosing 19 patients with varying stages of dementia and 10 patients showing other disorders except dementia. EVINCE-I and the human expert were in perfect agreement on the diagnosis of dementia and correlated highly on the diagnosis of dementia of the Alzheimer type and multiple infarct dementia. EVINCE-I thus offers important possibilities as a tool in investigating the data and procedures used by the human expert.

An expert system (ES) is a computer program designed for specific fields of expertise in which it attains a performance equal to or better than that of a human expert. More specifically, medical ESs can be regarded as computer software products with a medical data base that are designed to assist physicians and medical personnel in diagnosis, therapy, and related tasks in medical care (13,121). As the use of medical ESs is expected to produce profound changes in health care (11;13;15), it is of utmost importance to know the potential of this new technology. For instance, it is of relevance to know whether medical ESs are able to make similar—or even better—diagnoses than a physician and whether medical ESs could be of value in improving health care. In addition, it is important to acquire information about the possible benefits and risks of this new technology (4).

The medical ESs that have been developed until now share some characteristics. First, they have been developed for areas in which expert knowledge is fairly well defined, such as CENTAUR for the interpretation of pulmonary tests (17,275) and AI/MM for renal physiology (17,273). Second, a common feature of most medical ESs is that the amount of data used is relatively small. Third, hardly any medical ES developed so

far has left the prototypical or demonstration stage, and there is a paucity of studies in which the medical ES is compared with the physician, i.e., "domain expert." Fourth, there has been a very rapid increase in the use of personal computers in medicine, but until now no reports have been published about their potential use for medical ESs. Thus, it is of interest to know whether it is possible to develop a medical ES for more complex problems such as those in areas where physiological, behavioral, and psychosocial data have to be combined, whether such a medical ES can be developed on a (cheap) personal computer, and whether such a medical ES yields conclusions that are similar to those of the physician/domain expert. This article applies these questions to the field of neuropsychiatry and the increasingly important topic of the differential diagnosis of dementia. In the first part, a description will be given of the development of EVINCE-I. The second part presents the assessment of the medical ES EVINCE-I versus the domain expert in diagnosing 29 patients.

DEVELOPMENT OF EVINCE-I

Differential Diagnosis of Demential Syndromes

Dementia is a concept that still has very controversial definitions, although most experts agree that it is better regarded as a syndrome than as a nosological entity. Some essential features of the demential syndrome are: memory impairment, intellectual deterioration, and changes in personality (8). Dementia can have many different causes — up to 50 are enumerated by C. D. Marsden (9). The two main causes of dementia are Alzheimer's disease (AD), with a prevalence of 39–50% of the total number of cases, and multi-infarct dementia (MID), with a prevalence ranging from 13–30%. This leaves approximately 20–48% accounted for by the remaining causes (5;9;10). These broad ranges can be explained by the difficulty in defining the concepts of AD and MID. Some researchers regard the overlap between AD and MID as a special category, usually referred to as Mix. Moreover, some include depression-induced dementia, i.e., pseudodementia (10), which has a prevalence of approximately 9%, whereas others exclude this condition from their figures.

Given this description, the task of the ES in making a diagnosis in the individual patient is threefold:

1. deciding on the diagnosis "dementia";
2. searching for neurological signs and other somatic indices; and
3. interpreting these signs to decide on the etiology of dementia.

This means that the ES should have knowledge of common disorders that present themselves as dementia, e.g., depression, and disorders that preclude the diagnosis of dementia, e.g., delirium. After stating the diagnosis as being dementia, the ES should be able to differentiate between the possible causes of dementia. Since dementia is caused in 52–80% of all cases by AD and MID, and because depression is a major disorder that can present itself as dementia, it was decided to center the domain around these three causes. The possibility of adding some or all of the other causes in a later version of the ES was left open; their choice would depend on the data used for diagnosing the three main causes. This decision was made to economize on the amount of data needed in the present version of the ES.

Beside making valid diagnoses, i.e., a significant agreement between the ES and the domain expert, it was deemed important that both the ES and the domain expert express their certainty about any diagnosis made. Such a requirement would enable

us to make a more precise comparison between the ES and the domain expert. Furthermore, the ES should be able to reproduce the domain expert's process of medical decision making (MDM) for later evaluation of the rules and data used. This would simplify comparison of the differences and similarities found.

The Choice of the ES Shell

The most common structure for ESs is a computer program consisting of a knowledge base, an inference engine to infer new facts with the help of the rule base and available data, and a user interface to regulate the communication between the user and the inference engine (7).

ES shells consist of the above-mentioned units, excluding the knowledge base, and one of the major decisions to be made in the development of an ES is the choice of the appropriate ES shell. This is because the shell determines the type of knowledge representation and the inferences possible. For the development of EVINCE-I, the expert system tool ACQUAINT was chosen (1;16). Knowledge in ACQUAINT is represented in frames, comparable to a data base structure, which makes it relatively easy to add or delete knowledge. Reasoning in ACQUAINT can be both hypothesis- and data-directed, i.e., forward and backward, and includes the use of certainty factors (1;2;18) as a measure of confidence. Rules can be used at different levels of reasoning, i.e., rules controlling rules, and are organized in easy readable IF-THEN statements. Furthermore, ACQUAINT is operational on the widespread standard IBM-compatible PC with 512 K-RAM and two floppy drives. Another important feature is that the end-user is supplied with a run-time copy of the ES and not the whole tool, which makes it very attractive financially.

Defining the Knowledge Base

After the initial preparations of forming a general picture of the domain, a start was made by further defining the problem. This was done by drawing a detailed picture of the MDM strategy used by the domain expert, a neuropsychiatrist who is a member of the Department of Neuropsychology and Psychobiology at the State University of Limburg. This expert was chosen because of the methodology that he used, which is based on international guidelines for the diagnosis of dementia as detailed in Diagnostic & Statistical Manual (DSM-III-R) (3) and the report of the NINCDS-ADRDA consensus work group (12). Thus, the knowledge to be implemented would be widely acceptable.

The domain expert was therefore interviewed about these guidelines, and the resulting protocols were used to construct a decision tree which reflects the search strategy used by the domain expert (Figure 1). The decision tree is organized into contexts, each of which controls a body of rules and data concerning the subject to be investigated. The data used by EVINCE-I can be found in Table 1 and the diagnoses in Table 2. The global decision procedure will be discussed next.

Description of the Decision Procedures

The root node, or top-level context, is called "Initiate" and contains an introduction to the system. The next context contains a few rules to decide whether the problem of investigation is part of the knowledge domain. If it fits the domain, or if the user is not sure, the actual consultation will begin.

The actual consultation begins by collecting general data about the patient, such as date of birth, age, sex, etc. in the context "Patient Data," and data from auxiliary investigations, such as blood and urine tests, in the context "Preliminary Data." In

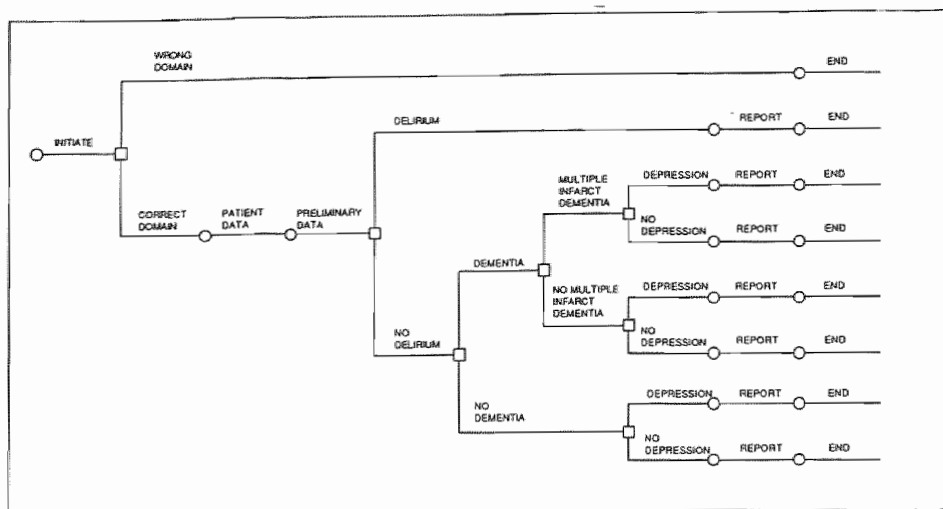


Figure 1. Context-level decision tree.

this latter context, EVINCE will check the data for possible deviations that could yield any preliminary diagnosis. These findings are then used during the examinations to follow. After that, EVINCE will try to find out the patient's state of consciousness. If the system is sure that the patient's consciousness is clouded, i.e., the symptoms fit the criteria for delirium, it will terminate the consultation. The system will then leap to the "Report" context and make a final report. The decision to abstain from further examination is taken on the basis that the consciousness should not be clouded, which is one of the DSM-III-R criteria for dementia (3).

If the patient's consciousness is not clouded, EVINCE will try to determine whether this particular case is an instance of dementia without making a decision about the etiology. EVINCE does so by using the remaining DSM-III-R (3) criteria for dementia. If dementia is not diagnosed, EVINCE-I will skip any context concerning dementia and jump to the context that determines whether depression is the cause of any of the problems encountered. In the "Report" context, it collects the diagnoses made thus far, checks them for mutual implications, and presents its findings.

If dementia is diagnosed, the system will try to determine the cause of the demential syndrome. Since AD is initially diagnosed by using exclusion criteria, the diagnosis AD is delayed until all other causes have been excluded. Therefore, EVINCE will first check the data for any vascular problems that could lead to the diagnosis of MID, the next major cause of dementia. Independent of the outcome, EVINCE-I will check the data for signs of depression. Finally, a last check is performed in the Report context to see if any of the preceding diagnoses can be used to exclude the diagnosis AD, e.g., a vitamin B₁₂ deficiency.

When all of these steps have been taken, and AD can still not be excluded, the ES will use some of the preliminary data to adjust the certainty factor of the final diagnosis AD. The certainty factor is raised, e.g., when the EEG pattern shows deceleration in the alpha-rhythm activity. The certainty factor is decreased if the EEG shows any signs of focal pathology. However, none of these data can exclude or confirm AD by itself; they are used as the NINCDS-ADRDA work group (12) proposes: for adjustment of the diagnostic confidence.

Table 1. Patient Characteristics Used by EVINCE-I

Subject	Contents
Identification data	Age, name, I.D. number, date of birth, consultation date, sex, hand preference, occupation, socioeconomic status, educational level
Hematology	BSE, Hb, Ht-percentage, leucocyte
Chemistry	Kalium, Na, Ca, P, Glucose, B ₁ , B ₁₂ , AF, Gamma-GT, OT, PT, T ₄ , TSH
Auxilliary tests	EEG, CT scan, ECG
Anamnesis	Weight, alcohol use (former and present), appetite, smoking habit (former and present), intoxication, occupational functioning, social functioning
Cognitive functions	Memory, abstract thinking, higher cortical functions (aphasia, apraxia, agnosia, constructive apraxia), consciousness, attention, judgment, orientation, perception
Motory	Speech, motor activity
Personality	Character, personality, social activities, disorder-insight
Complaint patterns	Onset, deterioration course, deterioration pace, physical complaints
Sleep	Somnolence, sleep-start disorder, sleep-through disorder, sleep-end disorder, nocturnal confusion
Mood	Lability, mood fluctuations, daytime fluctuations, guilt feelings, suicidal thoughts, inhibitions, agitation, fears with psychological or somatic manifestations (general and gastrointestinal), libido, hypochondria, depersonalization and depersonalization type, paranoid symptoms, symptoms of compulsion, Hamilton depression score
Atherosclerosis	Anamnestic hypertension, TIA, CVA, atherosclerosis, focal neurological symptoms, focal neurological signs, Hachinski score

Note: The list above does not include the additional variables and functions to monitor the search process of EVINCE-I, since they merely deal indirectly with the patient data and are system dependent.

Table 2. Diagnostic Knowledge Used by EVINCE-I

Demential syndrome (senile/presenile)
Primary degenerative dementia (senile/presenile)
Multiple infarct dementia
Major depression
Delirium
Cognitive disorders
Hyperthyreosis and hypothyreosis
Diabetes
Alcoholism
Electrolyte disorder
Inflammation
Vitamin B insufficiency
Medication
Epilepsy

The final report that EVINCE-I produces is made in the Report context. This context reflects the evaluative activity of the domain expert and the reformulation of the findings in legible standard phrases. This context, or child context, will write complete standard reports in future versions.

Table 3. Diagnostic Categories

Category	Diagnosis			
	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

Legend: DEM, dementia; DAT, dementia of the Alzheimer type; MID, multiple infarct dementia; DEP, depression; x, present, —, absent.

Table 4. Interrater Agreement per Diagnostic Category

Test Categories	Obs.	Exp.	Kappa	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

Note: All z scores are significant: $p < 0.0001$.
See Table 3 for the meaning of the categories.

COMPARISON BETWEEN EVINCE-I AND THE DOMAIN EXPERT

As stated earlier, it was deemed important to know whether a desktop ES is powerful enough to make valid diagnoses. Therefore, the performance of EVINCE-I was compared to that of the domain expert in an experiment in which the following two hypotheses were tested: (a) expert and EVINCE-I agree on their diagnoses; and (b) expert and EVINCE-I agree on the relative certainties of these diagnoses. Furthermore, the number of false-positive and false-negative diagnoses made by EVINCE-I was compared to that of the expert, in order to see whether a specific type of mistake prevailed in any diagnostic category.

Methods

From the patient records available from the Maastricht Memory Clinic, 19 patients were drawn who had been diagnosed with at least 50% certainty as having dementia, according to the estimation of the domain expert, irrespective of its cause or other diagnoses. These cases were taken to see whether EVINCE-I would make any false-negative dementia diagnoses. To determine whether EVINCE-I would make any false-positive diagnoses of dementia, 10 other patients were drawn who were not diagnosed as having dementia. Most of these patients suffered clinically from mild depression. The mean age of the first group was 73 (SD = 7.5) and that of the second group was 47 (SD = 14.6). None of the applied diagnostic criteria uses age as a distinguishing factor; it is reported here for the sake of completeness. All patients had been diagnosed by one domain expert, and none of the cases had been used in test runs when the system was being developed.

The expert received a form with a short instruction that asked him to record the

Table 5. Interrater Agreement Using an 11-point Certainty Scale

Category	Spearman's Correlate	df	t	Probability
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	$p < 0.002$

Note: $N = 29$; two-sided.

diagnosis of each patient in key words with a certainty value. This value could vary between (0) *unknown* to (100) *absolutely certain*. This certainty scale was used because EVINCE-I gave certainties within the same range. The data of the 29 patients given to EVINCE-I were entered by the knowledge engineer. All data requested by EVINCE-I were present in the patient records.

Results

The main diagnoses were the following: dementia (DEM), dementia of the Alzheimer type (DAT), multiple infarct dementia (MID), depression (DEP), and a residual category (RES). Since the RES category contained too great a variety of diagnoses for a meaningful comparison, this category was dropped from the statistical analysis. Each patient received one or more of the diagnoses. The categories of possible diagnosis made by the human expert and the system can be found in Table 3.

The first test examined the extent to which EVINCE-I and the domain expert were in agreement for all combinations of diagnosis made (the categories 1-8 in Table 3), in terms of the kappa value (14). The results can be seen in Table 4, test 1.

The overall level of agreement between the two raters was moderate. The reason that the overall agreement was not high can be found in the categories (7 + 8) and (3 + 4), which were supposed to make a distinction between patients diagnosed as depressed or not (see Table 3), but failed to do so adequately. The amount of agreement between the raters increased when the diagnosis of DEP was not taken into consideration, as can be seen in test 2 of Table 4.

In test 3, we analyzed the agreement between the two raters on the diagnosis of DEM, i.e., categories (1 + 2 + 3 + 4 + 5 + 6) and (7 + 8). This is an important diagnosis to test for, since the diagnosis of DAT and MID are dependent on it. Table 4, test 3, shows that the raters showed perfect agreement on this diagnosis. Since MID and DAT are considered to be mutually exclusive diagnoses in this system, we tested the agreement for the diagnosis DEM with DAT, i.e., categories (1 + 2), (3 + 4 + 5 + 6) and (7 + 8), and DEM with MID, i.e., categories (1 + 2 + 5 + 6), (3 + 4), and (7 + 8), separately. In both cases, the raters showed a high degree of agreement on their diagnoses (see Table 4, tests 4 and 5).

The results indicate that the agreement between EVINCE-I and the domain expert is "high" to "perfect" when corrected for the diagnosis DEP. Clearly, depression is the only diagnosis that is not handled correctly by EVINCE-I.

After this analysis, the ratings were transformed, with normal rounding, to an 11-point (0-10) scale in order to test whether the ratings agree in direction and relative magnitude. We used the same diagnoses but examined them separately since we were interested in only the individual diagnostic ratings. For each diagnosis we computed the Spearman rank correlation coefficient.

Table 5 shows that the raters displayed a high degree of correlation in their cer-

tainty about the diagnostic subjects, except for the diagnosis DEP, which showed a moderate correlation.

Although the RES category was dropped from the statistical analysis, we can still find some clues to the ability of EVINCE-I to diagnose rarer causes of dementia. For example, one patient was diagnosed by EVINCE-I as suffering from MID, but in fact had multiple sclerosis (MS). Since the system knew nothing about MS, MID is a logical near miss. Another patient's diagnosis contained the comment "cave vitamin B₁₂ deficiency" from the domain expert. EVINCE-I picked this possible cause up and made the same note in its report.

In two cases, the psychiatrist noted a deviation in the function of the thyroid gland, and EVINCE-I managed to note one of these two cases. Alcohol as a probable cause for dementia was noted in two cases by both the expert and EVINCE-I.

Discussion

Given the results presented, the tests confirm our notion that EVINCE-I can be considered a reasonably good replication of human expertise on the subject of dementia. The results further show that an ES makes it easier to trace problems in the domain knowledge. Furthermore, the fact that EVINCE-I uses no more than 122 concepts, which is substantially less than the total data gathered by the human expert, and still is able to produce these results, shows that an ES can be used as a means to assess the contribution of the data used by the expert. Further development of EVINCE-I and a test of these possibilities is currently being performed.

The fact that the correlation between the domain expert and EVINCE-I was high for the categories DEM, DAT, and MID provides evidence that the knowledge represented in EVINCE-I is sufficiently sophisticated to yield valid diagnoses within these three categories. EVINCE-I also seemed to do a good job for the more rarely occurring causes of dementia, when it had the knowledge to detect them. The only category it performed less well on is that of depression. In an examination of the reasons for the discrepancy between EVINCE-I and the domain expert, two possible causes are relevant. First, EVINCE-I might be relying too heavily on the interpretation of the Hamilton Depression Rating Scale (HDRS) (6). In an evaluative discussion, the domain expert stated that the HDRS-score is used as a measure of severity and not as a diagnostic criterion. Second, another cause for the low correlation may be due to a rule in EVINCE-I that decides whether to investigate depression. It does so only if a patient *appears* to be, or feel, depressed. This might be the pitfall for patients who do not show obvious signs of depression.

These causes explain why EVINCE-I made 4 false-negative and 2 false-positive diagnoses in this category. The first cause indicates that the system does not have sufficient criteria, while the second cause indicates that the system rejects investigating the possibility of depression too often. However, there is yet a further possibility, which is typical for the problems in this domain. Since a psychiatric diagnosis (especially depression) can hardly ever be determined with absolute certainty, it could be that the domain expert, and not EVINCE-I, made the wrong diagnoses. An experiment involving several human experts and an improved version of EVINCE has been performed to examine this possibility. The data of that experiment are being analyzed now.

Furthermore, the domain currently covered by EVINCE-I is expanded to include amnesic and dysthymic disorders. This could lead to a system that provides a standard protocol for clinicians and for researchers who need to select patients with dementia.

CONCLUSION

As stated in the introduction, the EVINCE-I project set out to assess the possibility of developing an ES for the differential diagnosis of demential syndromes. We further wanted to assess the possibility of whether such an ES could be implemented on a personal computer and whether it would yield conclusions that are similar to those of the physician. The results of the assessment presented above strongly confirm our belief that the use of such an ES is feasible and offers promising leads for use in future assessments. An ES on the subject of demential syndromes could, e.g., be a tremendous help in (a) mobilizing expert knowledge and making this knowledge accessible; (b) assessing the contribution of data to the reliability of the diagnosis; (c) examining the efficiency of the organization, procedures, and decisions in gathering data (i.e., time required, costs involved, inconvenience for the patient, etc.); and (d) investigating the effect of missing data when rediagnosing patients.

An ES used as a model system could thus offer not only the opportunity to assist in making reliable diagnoses but also provide a new technique to assess a broad range of subjects in the domain of expertise. At the moment, an adapted and expanded version of the ES, EVINCE-II, is being used to assess the topics mentioned earlier.

REFERENCES

1. ACQUAINT: *User's manual for Acquaint and Acquaint-Light*. Manual. Purmerend, Lithp Systems BV, 1987.
2. Buchanan, B. G., & Shortliffe, E. H. (eds.). *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*. Reading, PA: Addison-Wesley, 1984.
3. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 3rd Edition, Revised. Washington, DC: American Psychiatric Association, 1987.
4. Gurstein, M. Social impacts of selected artificial intelligence applications—the Canadian context. *Futures*, 1985, 652–71.
5. Haase, G. R. Diseases presenting as dementia. In C. E. Wells (ed.), *Dementia*. 2nd ed. Philadelphia, PA: Davis Co., 1971, 27–67.
6. Hamilton, M. Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, 1967, 6, 278–96.
7. Hayes-Roth, F., Waterman, D. A., & Lenat, D. *Building expert systems*, Reading, PA: Addison-Wesley, 1983, 210–12.
8. Jolles, J. Cognitive, emotional and behavioral dysfunctions in aging and dementia. In D. F. Swaab, E. Fliers, M. Mirmiran, et al. (eds.), *Progress in brain research*. Amsterdam: Elsevier Science Publishers BV, 1986, 70, 15–39.
9. Marsden, C. D. Neurological Causes of Dementia other than Alzheimer's Disease. In Kay & Burrows (eds.), *Handbook of studies on psychiatry and old age*, Amsterdam: Elsevier Science Publishers BV, 1984, 145–67.
10. Marsden, C. D. Assessment of dementia. In J. A. M. Frederiks (ed.), *Handbook of clinical neurology*, vol. 2. *Neurobehavioural disorders*. Amsterdam: Elsevier Science Publishers BV, 1985, 46.
11. Maxmen, J. S. Long-term trends in health care: The post-physician era reconsidered. In D. Schwefel (ed.), *Indicators and trends in health and health care*, Berlin: Springer Verlag, 1987.
12. 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.
13. Potthoff, P., Rothmund, M. Schwefel, D., et al. Expert Systems in Medicine. *International Journal of Technology Assessment in Health Care*, 1988, 4, 121–33.
14. Schouten, J. A. *Statistical measurement of interobserver agreement*. Unpublished doctoral thesis, Utrecht: 1985.

15. Schwartz, W. B. Medicine and the computer: The promise and problems of change. *New England Journal of Medicine*, 1970, 283, 1257-64.
16. Tello, E. R. ACQUAINT. In P. Lemmons (ed.), *BYTE*, Peterborough, NH: McGraw-Hill Inc., 1987, 7, 265-72.
17. Waterman, D. A. *A guide to expert systems*. Reading, PA: Addison-Wesley, 1986, 272-88.
18. Winston, P. H. *Artificial intelligence*, 2nd ed. Reading, PA: Addison-Wesley, 1984.