Differential diagnosis of dementia

Differential Diagnosis of Dementia: A Comparison Between the Expert System EVINCE and Clinicians

Leonard A. Plugge, M.Sc.
Frans R. J. Verhey, M.D.
Jellemer Jolles, Ph.D.

The diagnostic performance of the expert system EVINCE was compared with that of 85 clinicians in diagnosing 10 patients suspected of suffering from dementia. A multidisciplinary expert committee provided a standard diagnosis as reference for comparison. The results showed that the syndrome and etiologic diagnoses made by EVINCE were in very close agreement with those of the expert committee and that the diagnostic performance of EVINCE was better than that of the average clinician. The present findings indicate that expert systems, especially those within the realm of complex multidimensional medical problems, could be a valuable aid in medical practice.

(The Journal of Neuropsychiatry and Clinical Neurosciences 1991; 3:398–404)

The diagnosis of dementia and dementing diseases is based on neurological and psychiatric findings, but it is usually made by physicians from one discipline. In a previous study, we showed that such a monodisciplinary approach had a significant effect on the type of diagnoses made.[1] We compared the diagnoses of neurologists, psychiatrists, nursing home physicians, general physicians, and psychologists, and found that neurologists made the diagnosis of Alzheimer's disease more frequently than clinicians of any of the other disciplines. Consistent with, for example, DSM-III-R guidelines, psychiatrists used the diagnosis of depression more often than did neurologists, while the other disciplines took a middle position.[2] Furthermore, psychiatrists and nursing home physicians more often made a syndrome diagnosis without an etiologic diagnosis than did neurologists.

These findings are consistent with the results reported by Hoffman[3] and by Verhey et al.,[4] all of whom found that a thorough multidisciplinary neuropsychiatric examination resulted in a therapeutically important alteration in the referring diagnosis (41% and 45%, respectively). The reemergence of neuropsychiatry is an important step toward a solution for such discipline-related diagnostic biases. However, there are more complicating factors in the diagnosis of dementia. Despite the fact that international criteria have been

Received July 30, 1990; revised January 8, 1991; accepted January 25, 1991. From the Department of Neuropsychology and Psychobiology, University of Limburg, the Netherlands. Address reprint requests to Dr. Jolles, Department of Neuropsychology and Psychobiology, University of Limburg, Box 616, 6200 MD Maastricht, The Netherlands.

developed to improve the consensus on the definitions of dementia, these criteria are only slowly being applied, if at all, in daily practice. Furthermore, there is a lack of general consensus on the selection and weighing of data, as well as on the examination procedure. For instance, in the NINCDS-ADRDA criteria, no consensus could be reached about the selection of the neuropsychological methods to be used.[5]

One way dementia diagnostics and classification can be improved is to use computer-controlled medical protocols to gather the data.[6] However, these programs are usually sophisticated data bases with a disadvantage: the data set requested by the program is always the same, irrespective of patient characteristics. In addition, these programs do not incorporate knowledge of data integration, detection of data inconsistencies, or data relevance.

In a review of computer-based decision aids, Morelli et al.[7] compared five prominent decision-making paradigms: data bank analysis, statistical pattern recognition, Bayesian analysis, the logical flow chart, and knowledge-based expert systems. In their conclusions these authors stated that

> the expert system approach appears to be the most promising. Its main strengths are 1) the ability to incorporate different kinds of knowledge into the decision-making process, 2) the ability to mimic the way humans reason about a problem, 3) the ability to explain and justify the system's conclusions. . . .[7] (p. 166)

Based on these considerations, the neuropsychiatric expert system EVINCE was developed. EVINCE is based on international rules and criteria for dementia diagnostics as described in DSM-III-R and proposed by the NINCDS-ADRDA work group.[2,5] A neuropsychiatrist (the domain expert)—the second author of this paper—provided the expertise in applying these rules and criteria. A first evaluation revealed that the diagnoses produced by EVINCE showed a high level of agreement with those made by the domain expert.[8]

After this first evaluation, EVINCE was developed further. In order to test this expanded version of EVINCE, an evaluation experiment was set up involving a multidisciplinary committee of three expert clinicians and 85 clinicians from five different disciplines. The multidisciplinary expert committee (MEC) provided diagnoses as a reference for comparison with the diagnoses of EVINCE and of the 85 clinicians. The hypothesis was that EVINCE would make more correct diagnoses (that is, the number of diagnoses that are in agreement with those of the MEC) than the average clinician taking part in the experiment.

METHODS

Subjects

The subjects were participants in a consensus meeting on the differential diagnosis of dementia, organized by the Dutch National Organization for Quality Assurance in Hospitals (CBO) in the Netherlands, in November 1988. Each of the 458 registered participants was asked to cooperate in an inquiry concerning the present state of dementia diagnostics. As the inquiry data were to be used in a more extensive study on the use of classification in dementia diagnostics, the inquiry was divided into two parts—one before and one after the consensus meeting. (The results of this study will be published elsewhere.) Of the 458 participants, 127 people handed in their first form, and, of these 127, 90 filled out and returned their second inquiry form.

Based on information provided by the participants for the registration office of the CBO, 85 respondents represented five disciplinary categories: 1) neurologists, 2) psychiatrists, 3) general physicians, 4) nursing home physicians, and 5) psychologists, leaving a residual category of five respondents that was dropped from the analysis. The data from the 85 respondents were used in this study.

Materials

Ten case descriptions were selected from the patient records of the Maastricht Memory Clinic. The cases were selected so that both classic and more complex cases were present, with different levels of severity of the cognitive—or memory—disorder.

Each case description was formulated in terms of a standard medical report and contained all information necessary to make a diagnosis according to standard research criteria recommended by DSM-III-R and the NINCDS-ADRDA work group.[2,5] This information was incorporated in the following paragraphs: 1) introduction, 2) past history, 3) anamnesis, 4) anamnesis as reported by a partner or by a close member of the family, 5) psychiatric and neurological history, 6) medical history, 7) medication data, 8) intoxication data, 9) psychosocial data, 10) daily functioning, 11) physical examination, 12) neurological examination, 13) psychiatric examination, 14) blood examination, 15) neuropsychological examination, 16) additional examination (e.g., CT-scan, chest x-ray, EEG, or ECG). To guarantee the patient's privacy, any information that might identify the patient was either changed or omitted. Each participant was asked to answer questions concerning his or her age, years of medical experience, discipline, nature of medical practice, and hours per week spent on differential diagnosis and classification of patients suspected of

suffering from dementia. Finally, after examining the case descriptions, the participants were asked to write down their diagnoses on the form.

Multidisciplinary Expert Committee

To establish a reference for comparison of the diagnoses, an independent multidisciplinary committee of three expert clinicians was established, consisting of a neurologist, a psychiatrist, and a psychologist. Each member of the MEC received the same 10 patient case descriptions and was asked to study the data and to formulate diagnoses. The MEC was then given the opportunity to discuss these diagnoses in a joint conference to reach a consensus on the final diagnoses. The MEC was asked to state the diagnoses at both the syndrome and the etiologic level. The conference lasted approximately 4 hours. The MEC reached a consensus for all patients, except for the etiologic diagnoses of patients 4 and 6 (see Table 1).

Classification of Diagnostic Judgment

The consensus meeting was also concerned with diagnostic terminology; the inquiry data were going to be used for qualitative analysis of agreement on terminology at a later stage. No instruction was given to the participants as to which terminology or classification they should use. The only instruction given was to include all relevant diagnoses in key words.

To perform a meaningful quantitative analysis of these diagnoses, we constructed a classification system according to the following principles: Diagnoses on the syndrome level and the etiologic level were coded separately. On the syndrome level, the possibilities included 1) dementia, 2) cognitive disturbances not termed dementia, 3) no cognitive disturbances, and 4) no statement about cognitive functioning. The etiologic level was classified according to the following causes: 1) primary neurodegenerative, 2) cerebrovascular, 3) neurological other than 2, 4) internal, such as endocrine and/or metabolism, 5) drug-induced, 6) depression-induced, 7) related to psychosocial factors. Although the term *depression* usually is used to mean a syndrome (for instance, in DSM-III-R), it is used here in an etiologic sense, i.e., as a possible cause for dementia or cognitive deterioration (cf., "depression-induced dementia"). This makes it possible to avoid the term *pseudodementia.* Thus, the diagnostic statements of the 85 clinicians, the MEC, and EVINCE were classified at the syndrome and etiologic level.

TABLE 1. Summary of diagnoses made by the Multidisciplinary Expert Committee (MEC) and the expert system Evince

			MEC		Evince	
Case	Sex	Age	Syndrome Diagnosis	Etiology	Syndrome Diagnosis	Etiology
1	F	74	Moderate dementia, depressive symptoms	Probable AD	Mild dementia, depression	AD
2	F	80	Moderate dementia	Probable AD, neuroleptic-induced parkinsonism	Moderate dementia	AD
3	M	78	Severe dementia	MID, neuroleptic-induced parkinsonism	Moderate dementia	MID
4	F	66	Mild dementia	Possible AD, or major depression	Mild dementia	Major depression (possibly medication–induced), bereavement
5	F	71	Slight cognitive deficit (no dementia)	History of CVA, adjustment disorder with depressive symptoms	Cognitive deficit (no dementia)	Vascular problems, dysthymic disorder
6	M	62	Slight cognitive deficit (no dementia)	History of TIA, or mood disorder	Cognitive deficit (no dementia)	Vascular problems
7	F	80	Severe dementia	Probable AD	Severe dementia	AD
8	F	86	Mild dementia with depression	MID	Mild dementia	MID
9	M	72	Mild dementia	Possible AD, MID, depression	Mild dementia	AD, depression
10	M	67	Moderate dementia	MID	Moderate dementia	MID

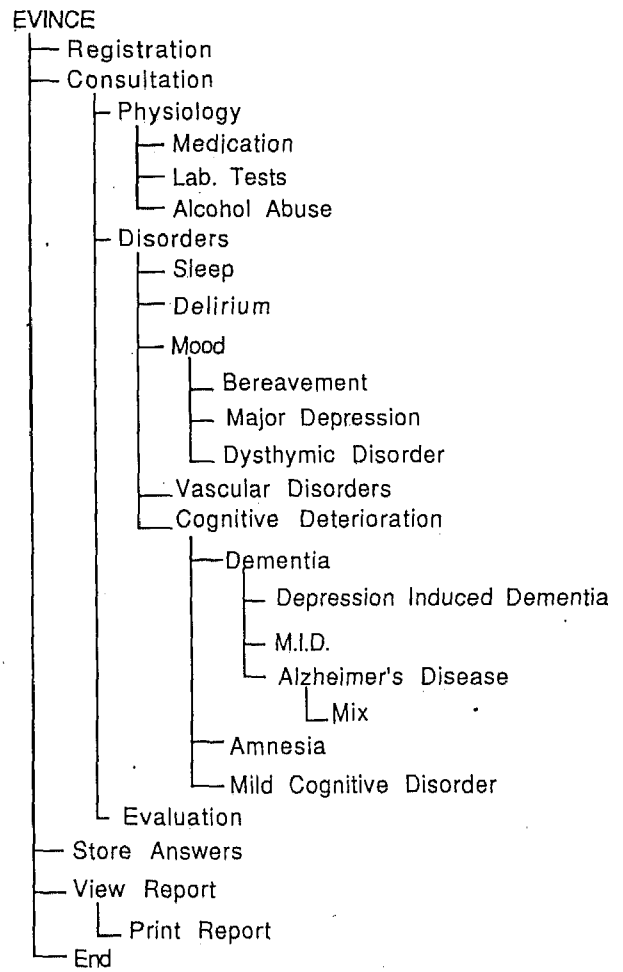
Note: AD=Alzheimer's disease; CVA=cerebrovascular accident; MID=Multi-infarct dementia; TIA=transient ischemic attack.

The Expert System EVINCE

EVINCE was developed with the expert system tool Acquaint.[9] The minimum requirements for EVINCE are as follows: an IBM PC-compatible microcomputer (preferably an AT) with a minimum of 520 Kbytes of RAM (640 Kbytes of RAM is recommended), two floppy drives (a hard-disk drive is recommended), and MS-DOS or PC-DOS Version 2.1 or later. EVINCE can be used with a monochrome or a color monitor.

EVINCE is a package consisting of the actual program, i.e., the inference engine and user interface, and three knowledge modules. In Module 1 the user can decide to consult EVINCE in batch mode (i.e., let EVINCE diagnose the preentered data of one or more patients) or in interactive mode (i.e., the system asks questions and the user provides the answers). Modules 2 and 3 consist of the procedural knowledge depicted in Figure 1. These two modules embody 28 contexts, 110 rules, and 129 concepts. Additionally, these two modules use 143 formulas, i.e., functions that perform calculations, window and file management, etc. Because the knowledge modules are separate from the actual program, they also can be stored separately (e.g., on a network) to prevent unauthorized use.

FIGURE 1. Context level decision tree



Knowledge in EVINCE is represented in rule and concept frames. Concept frames represent the knowledge of patient data, while the rules embody the procedural knowledge. When the value of a concept is unknown to the system and requested by a rule, it either will be inferred whenever possible or requested from the user. A rule consists of an IF and a THEN part. The IF part compares the concept value with the test value and triggers the THEN part of the rule when the values match. Each concept value has a certainty attached to it; for example "maybe" = 50, and "unknown" = 0. These values range between −100 (absolutely false) and 100 (absolutely true) and are used to assess with how much certainty a conclusion can be drawn. Additionally, each conclusion has a "certainty" value that informs the system of how certain that conclusion is when the premises are absolutely true. Consequently, the lower the certainty of the premises, the lower the certainty of the conclusion.

As stated in the introduction, EVINCE was developed further on the basis of the results of a previous experiment. The knowledge base was extended to include knowledge about disorders that can be related to medication.[10] The diagnostic capabilities on the subject of depression were refined with DSM-III-R criteria and were extended to include the diagnosis "dysthymic disorder." The knowledge about dementia was reorganized to investigate the level of deterioration, which then would lead to the diagnosis of dementia or amnesic syndrome. This resulted in a hierarchical examination protocol consisting of diagnostic topics called "contexts" (Figure 1).

Each context governs a set of rules and sometimes one or more subcontexts (e.g., child contexts). A context is in fact a higher order rule that establishes whether it is worth checking its subordinate rules and contexts. If the system considers a context irrelevant, then all the subordinate rules and contexts also will be considered irrelevant. For example, the context "Mood" tries to establish whether there is reason to assume that the patient is depressed. If this is not the case, then all child contexts will be ignored. However, some contexts always will be examined because they concern information that is considered relevant (i.e., the contexts "Registration," "Medication," and "Lab-Tests"). Another context that always is used is the "Evaluation" context. In this context, all the diagnostic information gathered from the previous contexts is collected, checked for inconsistencies and mutual

consequences, and finally transformed into natural language statements. These statements then are printed in the form of a report. The last context is called "End" and gives the user the opportunity to ask the system how the presented conclusions were reached. It should be noted that such questions also can be asked during the actual (interactive) consultation.

RESULTS

Table 2 shows the number and percentage of respondents compared with those present during the consensus meeting. None of the disciplines of the respondents was over- or underrepresented in comparison with the number of participants per discipline.

The MEC reached a consensus on the syndrome diagnosis for all 10 patients. However, the MEC did not reach a consensus on the etiologic diagnosis for patients 4 and 6. Two etiologic diagnoses were given for these patients (Table 1). A diagnosis made by the clinicians or EVINCE was considered "correct" if the diagnosis was made by the MEC (Table 1). Thus, 10 points could be scored for the syndrome, and 10 points for the etiologic diagnoses. The diagnoses were not compared for the level of severity of the demential syndrome, or for the use of the classifications "possible" and "probable" for Alzheimer's disease because these frequently were omitted by the 85 clinicians. Comparison of the syndrome diagnoses of the 85 clinicians with the MEC revealed that they had a mean±SD of 7.6±1.4 correct diagnoses (n=85; range 5–10). No differences were found between the disciplines (Table 3).

EVINCE diagnosed all 10 case descriptions correctly, i.e., in agreement with the MEC. However, when the level of severity of the demential syndrome was taken into account, it was revealed that EVINCE considered patients 1 and 3 to be less severely demented than did the MEC (i.e., mild vs. moderate). Furthermore, EVINCE did not provide a level of severity for cognitive deficits (patients 5 and 6) because it was not designed to do so. (See the two right columns of Table 1.) On the etiologic level, the 85 clinicians reached a mean of 5.3±1.7 correct diagnoses (n=85; range 1–8). EVINCE, however, made all 10 etiologic diagnoses in agreement with the expert committee (Table 1). However, EVINCE did not use the classification "possible" and "probable," as did the expert committee. (The present new version incorporates rules to resolve this problem.) Furthermore, in contrast to the MEC panel, EVINCE did not make the diagnosis multi-infarct dementia for patient 9. The MEC decided that this diagnosis should not be excluded completely, given the finding of a small hypodensity on the CT-scan. In accordance with international consensus, EVINCE did not consider the CT-scan finding alone sufficient to make the diagnosis multi-infarct dementia, since the patient's history and examination did not reveal a cerebrovascular accident.

With respect to the etiologic diagnoses, a significant difference was found between the disciplines concerning the number of etiologic diagnoses that agreed with the expert committee. Neurologists had significantly more correct etiologic diagnoses than clinicians of the other disciplines, except for psychologists (see Table 4).

DISCUSSION

The results show that the average clinician made fewer etiologic diagnoses than syndrome diagnoses that were in agreement with the MEC. This can be explained partly by the fact that there are fewer choices in the latter. Thus, it is possible to have a lower score for the etiologic diagnoses by chance. However, the diagnostic performance of EVINCE was not affected by these differences in chance. Furthermore, the average score of the clinicians for both syndrome and etiologic diagnoses was considerably lower than the score of EVINCE.

Another important finding is the disciplinary difference found in the number of diagnoses that were in agreement with the MEC, specifically the difference be-

TABLE 2. Number of participants and respondents in absolute figures and percentage of total

Discipline Category	No. of Participants (%)	No. of Respondents (%)
Neurologists	100 (23)	24 (28)
Psychiatrists	57 (13)	13 (15)
General physicians	74 (17)	12 (14)
Nursing home physicians	133 (31)	26 (31)
Psychologists	66 (16)	10 (12)
Others[a]	28 (6)	5 (6)
Total	**458 (100)**	**90 (100)**

[a]The residual category Others was dropped from the analyses, leaving a total of 85 respondents.

TABLE 3. Mean number of syndrome diagnoses in agreement with the MEC

Discipline Category	n	Mean±SD	Range
Neurologists	24	7.833±1.31	5–10
General physicians	12	7.667±1.37	5–10
Nursing home physicians	26	7.654±1.33	5–10
Psychologists	10	7.300±1.83	5–10
Psychiatrists	13	7.100±1.44	5–9

tween psychiatrists and neurologists. Both neurologists and psychologists had significantly more etiologic diagnoses in agreement with the MEC than did psychiatrists. Our previous study[1] on interdisciplinary differences revealed that psychiatrists more often made a syndrome diagnosis without an etiologic diagnosis, in contrast to the neurologists. This difference between neurologists and psychiatrists was also seen in the present study. The possibility that this difference is due to the level of experience and/or involvement in dementia diagnostics could not be established. There was no significant difference between disciplines concerning health care experience. A significant difference was found concerning the time spent on dementia diagnostics (χ^2 = 13.76, df=5, p<0.008); psychologists and general physicians spent more time on dementia diagnostics than the other disciplines. However, this does not explain the discrepancy between neurologists and psychiatrists. A more plausible cause of interdisciplinary difference would be the fact that the 85 clinicians were not given the opportunity to discuss the cases in order to develop a multidisciplinary consensus. However, this monodisciplinary approach does not deviate from what is common practice at present. The question as to whether the difference between neurologists and psychiatrists might be due to the nature of the medical specialization and to the clinicians' experience is the subject of a forthcoming study.

Although most clinicians are familiar with written case reports as an alternative to seeing the patient in person, it is possible that this has had a negative influence on their results. However, it should be noted that the standardized patient information was given to both the 85 clinicians and the MEC panel. As Lopez et al. have remarked, this ensures a reduction of variance stemming from the patients and the clinicians.[11] Another possible cause for the low etiologic performance of the clinicians might be the rather low response rate. Although 85 (67%) of 127 clinicians responded to both the first and the second inquiry, they account for only 18% of the total number of participants. However, it should be noted that the participants who responded both times considered themselves competent, while others returned an empty form with the remark that they were too inexperienced.

Although the diagnoses made by the MEC could not be compared with postmortem and/or long-term follow-up data, it is thought that these diagnoses were reliable because the three clinicians involved were recognized experts in their discipline, and they had ample opportunity to discuss each case thoroughly. As mentioned in the introduction, the overall results of this experiment are in agreement with those described by Hoffman[3] and Verhey et al.,[4] who assessed referral diagnoses of behavioral disorders with a multidisciplinary team.

The aim of comparing EVINCE with human clinicians was to assess the performance of the system, not to show that an expert system can replace the human clinician. Although expert systems are able to mimic human reasoning, they (still) lack important human capabilities (such as intuition), as well as the vast amount of world knowledge. However, the results warrant the conclusion that the implemented neurological, psychiatric, and psychological knowledge was successfully applied by EVINCE with the material presented and that the system can assist clinicians in diagnosing dementia. As Hoffman observed: "It is clear that the techniques of neuropsychiatric diagnosis have currently advanced to the point where a major limitation exists in their knowledge of application. . . ." (p. 967).[3] However, as Teitelbaum noted, we cannot expect clinicians to be experts in all disciplines.[12] Nevertheless, they should be able to recognize situations that demand referral and collaboration. A computer-based decision aid like EVINCE could help to achieve this. It provides the clinician the opportunity to apply up-to-date knowledge of internationally approved standard criteria, and it can help to make the clinician aware of the decision process involved and provide a useful check for completeness of the necessary information. Furthermore, it offers the clinician a tool to integrate expert diagnostic information from different disciplines.

Although it is unlikely that expert systems will replace the human clinician due to the limitations noted earlier, it is to be expected that, in the near future, expert systems will play an important role in assisting medical diagnostics.[13,14] However, before expert systems are given such an important function in medical diagnostics, their performance should be thoroughly tested, both in laboratory and in field situations. Therefore, collaborative studies have been initiated to compare the diagnoses made by EVINCE both retrospectively and prospectively with clinical and postmortem diagnoses, as well as to test the

TABLE 4. Mean number of etiological diagnoses in agreement with the MEC

Discipline Category	n	Mean±SD	Waller Grouping		
Neurologists	24	6.208±1.38	A		
Psychologists	10	5.900±1.52	A	B	
General physicians	12	4.917±1.38		B	C
Nursing home physicians	26	4.885±1.68		B	C
Psychiatrists	13	4.615±1.94			C

Note: The Waller Grouping k ratio t test makes it possible to compare means of several groups at the same time, while minimizing the Bayes risk under additive loss. This means that a correction is made, using Bayes theorem, for the increased chance of finding significant differences due to multiple comparisons. k ratio=100, df=80, MSE=2.518, F=3.53, t=2.11. (Means with the same letter were not significantly different.)

system in field situations. Finally, the authors welcome other proposals for collaborative studies, particularly from outside the Netherlands.

The authors express their gratitude to Prof. Dr. B.P.M. Schulte (chairman of the CBO Work Group Dementia), Dr. J.J.E. van Everdingen (CBO staff member), Dr. M.F.A. Diesfeldt, Dr. J.A.M. Frederiks, and Prof. Dr. W. van Tilburg for their indispensable cooperation during the study.

This research was supported by the Medical Technology Assessment Program of the Dutch Ministry of Welfare, Public Health and Culture (W.V.C.) (TA-87-19) (CRO-236509).

References

1. Plugge LA, Verhey FRJ, Van Everdingen JJE, et al: Differential diagnosis of dementia: an experimental study into intra- and inter-discipline agreement. J Geriatr Psychiatry Neurol 1991; 4(2):90–97
2. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 3rd Edition, Revised. Washington, DC, American Psychiatric Association, 1987, p 106
3. Hoffman RS: Diagnostic errors in the evaluation of behavioral disorders. JAMA 1982; 248(8):964–967
4. Verhey FRJ, Vreeling FW, Jolles J: DSM-III and NINCDS/ADRDA criteria for dementia and Alzheimer's disease: impact of diagnostic procedures on daily practice, in Alzheimer's Disease: Proceedings of the Fifth Meeting of the International Study Group on the Pharmacology of Memory Disorders Associated With Aging, Zürich, January, 1989. Edited by Wurtman J. Cambridge, MA, Center for Brain Sciences and Metabolism Charitable Trust, 1989, pp 419–423
5. McKhann G, Drachman D, Folstein M, et al: Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984; 34:939–944
6. Roth M, Tym E, Mountjoy CQ, et al: CAMDEX: A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. Br J Psychiatry 1986; 149:698–709
7. Morelli RA, Bronzino JD, Goethe JW: Expert systems in psychiatry: a review. J Med Syst 1987; 11(2–3):157–168
8. Plugge LA, Verhey FRJ, Jolles J: A desk-top expert system for the differential diagnosis of dementia: an evaluation study. International Journal of Technology Assessment in Health Care 1990; 6(1):147–156
9. ACQUAINT: User's manual for Acquaint and Acquaint-Light Manual. Purmerend, the Netherlands, Lithp Systems BV, 1987
10. World Health Organization: Drugs for the Elderly. Copenhagen, World Health Organization Regional Office for Europe, 1985
11. Lopez OL, Swihart AA, Becker JT, et al: Reliability of NINCDS-ADRDA clinical criteria for the diagnosis of Alzheimer's disease. Neurology 1990; 40:1517–1522
12. Teitelbaum ML: Toward better integration of medical and psychiatric care. JAMA 1982; 248(8):977
13. Maxmen JS: Long-term trends in health care: the post physician era reconsidered, in Indicators and Trends in Health and Health Care. Edited by Schwefel D. Heidelberg, Springer-Verlag, 1987, pp 109–115
14. Potthoff P, Rothemund M, Schwefel D, et al: Expert systems in medicine: possible future effects. International Journal of Technology Assessment in Health Care 1988; 4:121–133