MRI of the Hippocampus in Alzheimer's Disease: Sensitivity, Specificity, and Analysis of the Incorrectly Classified Subjects

ARTICLE *in* NEUROBIOLOGY OF AGING · JANUARY 1998

Impact Factor: 5.01 · DOI: 10.1016/S0197-4580(98)00006-2 · Source: PubMed

9 AUTHORS, INCLUDING:

Mikko P Laakso
Niuvanniemi Hospital
94 PUBLICATIONS **7,305** CITATIONS
SEE PROFILE

Hilkka Soininen
University of Eastern Finland
827 PUBLICATIONS **29,449** CITATIONS
SEE PROFILE

Tuomo Hänninen
Kuopio University Hospital
129 PUBLICATIONS **6,869** CITATIONS
SEE PROFILE

Pauli Vainio
Kuopio University Hospital
120 PUBLICATIONS **6,256** CITATIONS
SEE PROFILE

ELSEVIER

Neurobiology of Aging, Vol. 19, No. 1, pp. 23–31, 1998

0197-4580/98 $19.00 + .00

PII:S0197-4580(98)00006-2

MRI of the Hippocampus in Alzheimer's Disease: Sensitivity, Specificity, and Analysis of the Incorrectly Classified Subjects

M. P. LAAKSO,*† H. SOININEN,[1]* K. PARTANEN,† M. LEHTOVIRTA,* M. HALLIKAINEN,* T. HÄNNINEN,* E.-L. HELKALA,* P. VAINIO,† AND P. J. RIEKKINEN SR*‡

*Department of Neurology, Kuopio University Hospital, Kuopio, Finland
†Department of Radiology, Kuopio University Hospital, Kuopio, Finland
‡A. I. Virtanen Institute, Kuopio, Finland

Received April 1, 1997; Revised November 18, 1997; Accepted November 26, 1997

LAAKSO, M. P., H. SOININEN, K. PARTANEN, M. LEHTOVIRTA, M. HALLIKAINEN, T. HÄNNINEN, E.-L. HELKALA, P. VAINIO, AND P. RIEKKINEN Sr. *MRI of the hippocampus in Alzheimer's disease: sensitivity, specificity, and analysis of the incorrectly classified subjects*. NEUROBIOL AGING **19**(1) 23–31, 1998.—In this study, magnetic resonance imaging (MRI) of the hippocampus for the diagnosis of early Alzheimer's disease (AD) is evaluated. We measured hippocampal volumes and the area of the medial hippocampus with a 1.5 T MR imager in 160 subjects: 55 patients with probable AD according to the NINCDS-ADRDA criteria, 43 subjects fulfilling the NIMH criteria of age-associated memory impairment (AAMI), 42 cognitively normal elderly controls, and 20 controls younger than 50 years. Three methods for normalization were compared. The hippocampi were atrophied in the AD patients, but not in the AAMI subjects or the elderly controls. There was no significant correlation between hippocampal volumes and age in the nondemented subjects. The discrimination based on volumetry resulted in an overall correct classification of 92% of AD patients vs. nondemented elderly subjects, whereas discrimination based on hippocampal area was less accurate, producing a correct classification in 80% of the subjects. We conclude that the hippocampus as assessed by MRI volumetry is atrophied early in AD, and spared by aging or AAMI. A brief critical review of previous studies is in concordance with the presented data: all the previous studies that have used volumetry, have similarly ended up with a good classification, whereas simpler or subjective measurements, subject to various sources of bias, have produced most variable results. © 1998 Elsevier Science Inc.

Age-associated memory impairment Aging Alzheimer's disease Dementia Diagnosis Hippocampus Magnetic resonance imaging

ALZHEIMER'S disease (AD) is the most common cause of dementia, yet difficult to diagnose precisely at the very onset of the disease. When novel pharmaceuticals to halt the progression of AD become available, the early diagnosis is of importance. The diagnostic process of cognitive dysfunction can be facilitated by magnetic resonance imaging (MRI), which has a superior contrast over other imaging methods. Combined with multiplanar observation possibility, MRI enables detection of changes in small, irregularly shaped structures, such as the hippocampus.

The hippocampus is part of the mesial temporal lobe memory system (50), and also known to be damaged and atrophied in AD at the very earliest stages of the disease (5,24). This finding has previously led to several MRI studies trying to evaluate hippocampal atrophy in AD, and its diagnostic use. Yet the results of these studies have varied vastly depending upon study setting, and, therefore, common agreement of usefulness of these measurements remains to be settled (9,13,17,27,29,30,35–37,40,41,48,56).

The purpose of this study was to further assess the MRI of the hippocampus to help establish the value of these measurements in AD diagnostics. In addition to AD patients, we examined two control groups of young and elderly cognitively intact individuals, and a group of non-demented subjects with age-associated memory impairment (AAMI). In brief, AAMI is a controversial concept consisting of people 50 years of age or older who suffer from objectively and subjectively detected memory impairment, but who are not demented (8,25). This study included 160 carefully screened subjects. Two methods for evaluation of hippocampal atrophy was compared: the accuracy of a simple measurement of hippocampal area from a single slice was compared to that of the measurement of the entire hippocampal volume. Because the current clinical criteria, no matter how careful the screening, have limited accuracy, we further profiled the subjects who were incorrectly classified according to MRI volumetry to assess the final accuracy of MRI measurements. Also, because no standard

[1]Address correspondence to: Hilkka Soininen, Department of Neurology, University of Kuopio, P.O. Box 1627, 70211 Kuopio, Finland; E-mail: Hilkka.Soininen@uku.fi; URL: http://www.uku.fi/laitokset/neuro/

TABLE 1
CLINICAL CHARACTERISTICS OF THE STUDY GROUPS

	AD	AAMI	OC	YC	ANOVA F
Number	55	44	42	20	–
Male/Female	28/27	11/32	19/23	10/10	–
Age, years	70 ± 8	70 ± 5	72 ± 4	28 ± 7‡	273*
Age range	48–83	51–78	64–79	21–43	
Education, years	7 ± 3	8 ± 3	10 ± 3‡	16 ± 3‡	47*
MMSE	22 ± 4†	28 ± 2	28 ± 1	–	77*

Results are expressed as mean age ± SD. ANOVA over the study groups: *$p < 0.0001$; Duncan post hoc analysis: †differs from AAMI and OC, $p < 0.05$; ‡differs from all other groups, $p < 0.05$.

for normalization exists either, we compared three simple methods for normalization of the data.

METHODS

Subjects

A total of 160 subjects were examined in the study: 55 patients fulfilling the NINCDS-ADRDA criteria of probable AD (38), 43 subjects fulfilling the National Institute of Mental Health (NIMH) criteria of AAMI (8), 42 cognitively normal age-matched controls, and 20 cognitively normal controls younger than 50 years of age. The clinical data of the study groups are presented in Table 1. The study was approved by the local ethics committee. All subjects provided their informed consent for participation in the study following an explanation of the study protocol.

The AD patients chosen for the study were recently diagnosed. They underwent the following evaluations: general physical and clinical neurological examination; assessment of clinical severity using Mini-Mental Status Examination (MMSE) (14), Clinical Dementia Rating scale (CDR) (23), and Brief Cognitive Rating Scale (BCRS) (44); comprehensive neuropsychologic testing; assessment of extrapyramidal signs using the Webster Parkinson's Disease scale (52); assessment of depressive signs by the Hamilton scale (19); an extensive battery of laboratory tests to exclude secondary causes of dementia; EEG and event-related potentials; and computed tomography (CT), single photon emission computed tomography (SPECT), and MRI of the brain. The AAMI group and the old controls (OC) were derived from a randomly selected population of 1049 individuals who participated in an epidemiological study on the prevalence of dementia and memory disorders (31). With minor exceptions, their examination battery was the same. The young controls (YC) were students or staff members volunteering for the study. They were healthy and, when interviewed, had no history of central nervous system (CNS) or systemic diseases. All the study subjects scored less than four in the modified ischemic scale (46).

Neuropsychological Tests

Verbal memory was examined with the list learning test using shopping items (20). A "yes" or "no" recognition of the words in the list was asked after a 30-min. delay filled with other psychometric tests. Boston approach was used to measure story recall (39). The recall of the story was tested immediately and after a 30-min. delay. Visual memory was examined with Heaton Visual Reproduction Test (47). The recall of the figures was tested both immediately and after a 30-min. delay. Buschke-Fuld Selective Reminding Test (7) was used to examine verbal memory in the groups of elderly controls and AAMI subjects. This test produced two variables: total recall and long-term memory. The total recall score comprised all the words the subject recalled from the list during six trials. Long-term memory score included all the words the subject recalled in two or more consecutive trials.

To assess executive functions, Trail-Making test A and B (45), and Verbal Fluency (4) were used. The maximum time of 150 s for Trail-Making A and 300 s of Trail Making B was allowed. If the test was not completed in the time allowed, the missing letters or numbers were scored as omissions. In the Verbal Fluency Test, the subject was asked to produce as many words as they could beginning with letters P, A, and S in 1 min. for each letter. The score was the number of words correctly named. We also used a variety of other tests assessing verbal, visuospatial, praxic, and executive functions (data not presented).

MRI Technique

The subjects were scanned with a 1.5 T Magnetom (Siemens, Erlangen) using a standard head coil and a tilted coronal 3D gradient echo sequence (MP-RAGE: TR 10 ms, TE 4 ms, TI 250 ms, flip angle 12°, FOV 250 mm, matrix 256 × 192, 1 acquisition). This resulted in T_1-weighted partitions with slice thickness of 1.5–2.0 mm, or approximately 20 slices per hippocampus, perpendicular to the long axis of the hippocampus. The slice thickness was randomly selected by the imaging personnel within the range. The MRI evaluations were performed by a single rater blinded to the clinical data or diagnostic category of the subjects except for age and gender.

The boundaries of the region of interest (ROI) were outlined by a trackball-driven cursor and the number of voxels within the region was calculated by using an in-house developed program for standard work console. The outlining of the boundaries always proceeded from anterior to posterior. The hippocampus included the dentate gyrus, the hippocampus proper and the subicular complex. The rostral end of the hippocampus was the anatomical beginning of the head when it first appears below the amygdala. The caudal end of the hippocampus was taken as the section in which the crura of the fornices departs from the lateral wall of the lateral ventricles. The methodology has been described elsewhere in more detail (35,49). The hippocampal area refers to the section of the hippocampus measured at the level of the anterior commissure. Examples on delineation of the hippocampus are shown in Figure 1.

In order to exclude individual size variability for the statistical analysis, we compared different simple methods for volume normalization. First, the volumes were normalized to coronal (ICA1) (Fig. 2) and sagittal (ICA2) (Fig. 3) intracranial areas, which account for all three dimensions of the cranium. The ICA1 was measured at the level where the anterior commissure first was

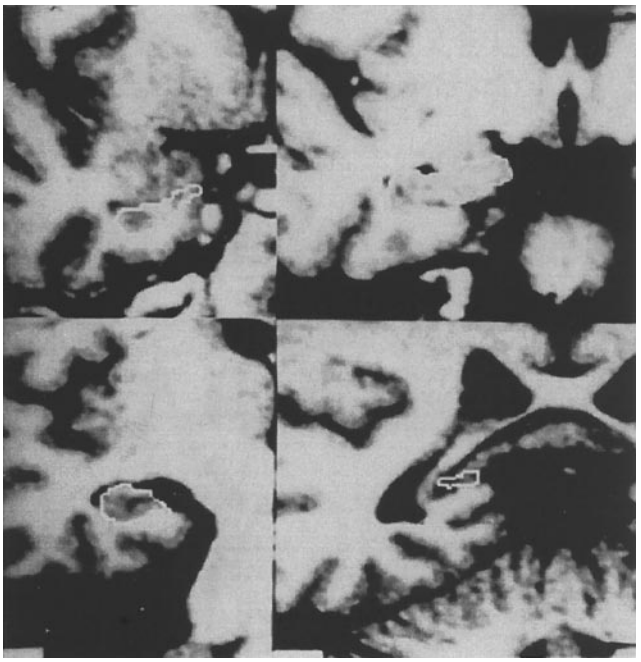

FIG. 1. Delineation of the hippocampus of an Alzheimer patient. The most posterior slice is on the lower right panel.

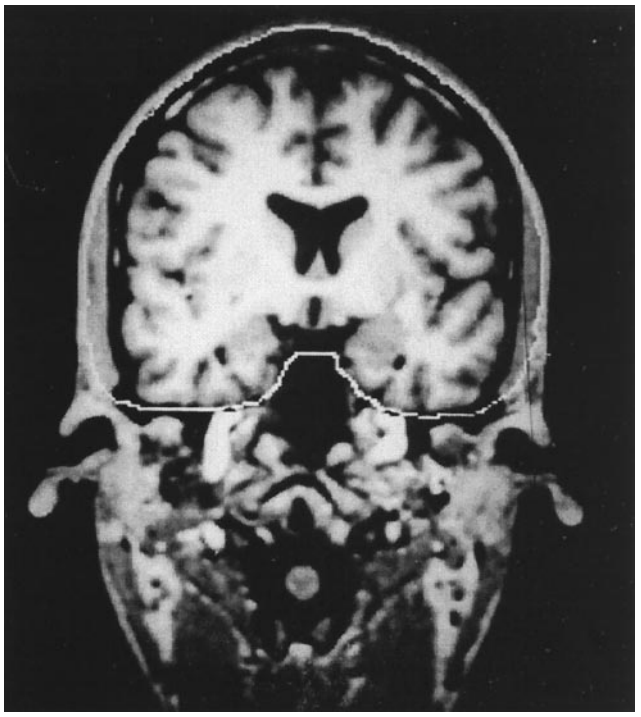

FIG. 2. The delineation of coronal intracranial area (ICA1).

present when proceeding from anterior, and the ICA2 was measured at the midsagittal scout image. The third normalization was done by multiplying hippocampal volumes by brain area/cranial area relationship. This was done in order to study the effect of general atrophy to the discrimination result. The brain area (Fig. 4) refers to the area of brain measured at the level of the anterior commissure with the gyral, lateral, and temporal ventricular spaces excluded. The volume of hippocampus for normalization was multiplied by 100 for improved readability, and also to emphasize its effect on brain area/ICA1 analysis.

The intrarater (test-retest) (49) and the interrater (35) agreements of the volumetry between two raters has been previously reported in detail. The intraclass correlation coefficient was 0.954 in the intrarater analysis. The differences between the volumes obtained by two raters compared to the mean of these two measurements were 4.1% for the right hippocampus and 1.6% for the left hippocampus.

Statistical Analysis

The data were analyzed by utilizing SPSS-PC+ V.4.1 software. Analysis of variance (ANOVA) with Duncan post hoc analysis was used to compare the means over the study groups. Correlations were calculated by using two-tailed Pearson's correlation test. To test the accuracy of the measurements to distinguish the AD patients from the controls we used stepwise discriminant function analysis (Wilks' method). To test whether the discriminant function gives similar discrimination in another sample of AD patients and controls, a random 50% sample was drawn from the study population to generate a discriminant function and this function was then tested in the other half of the study population. The results are expressed as mean $\pm$ standard deviation (S.D.). The level of statistical significance of differences is $p < 0.05$.

Analysis of the Incorrectly Classified

For this analysis, the study subjects were rearranged into four groups based on the results of the discrimination function analysis that included hippocampal volumes: AD patients correctly classified (ADCC), AD incorrectly classified (misclassified; ADMC),

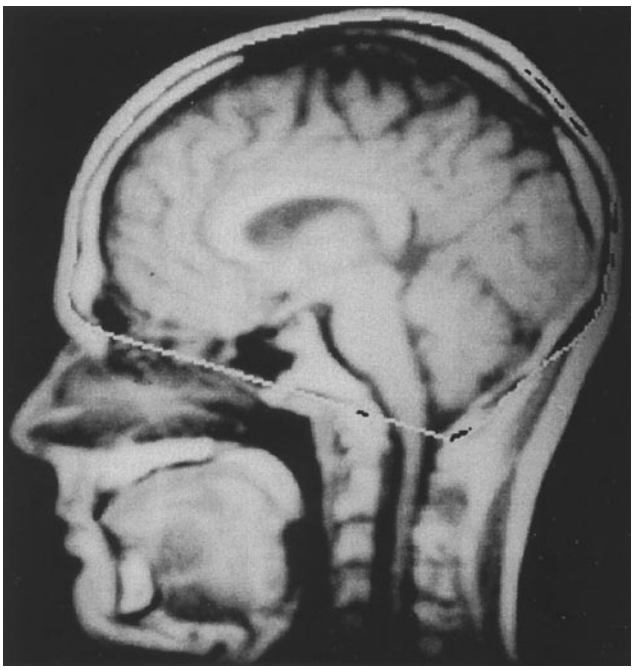

FIG. 3. The delineation of sagittal intracranial area (ICA2).

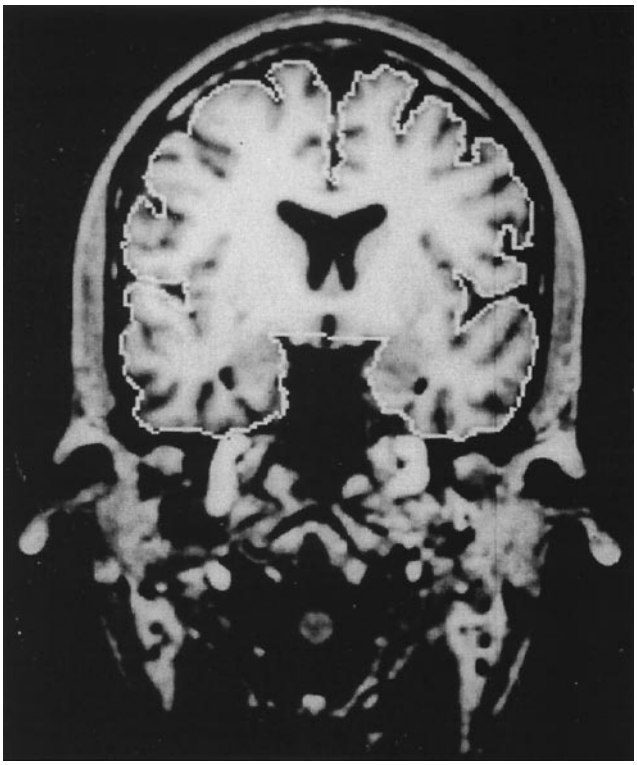

FIG. 4. The delineation of brain area. The lateral and temporal ventricular spaces were further excluded.

controls correctly (CCC), and incorrectly classified (CMC). The control group includes only the AAMI and the OC groups. Thus, and due to some missing data, the number of subjects varies from 137 to 140. The size of the individual study groups varies as follows: ADCC 46–49, ADMC 6–9, CCC 76–79, CMC 6–9. Demographic and clinical data such as disease severity was evaluated individually and within the group.

RESULTS

The AD patients and older controls did not differ significantly in age (Table 1). In the AAMI group, women were significantly overrepresented ($p < 0.05$). The controls had had more education than the AD patients (ANOVA, $F(3,157) = 47$, $p < 0.0001$). As expected, the MMSE scores were lower for the AD patients than for age-matched study groups($F(1,138) = 77$, $p < 0.0001$). According to the CDR scale, 5 AD patients had questionable (0.5), 35 had mild (1), and 15 had moderate (2) dementia.

Hippocampal Volumetry

Table 2 presents the mean hippocampal volumes for all the study groups and separately for men and women. AD patients had significantly smaller volumes of both hippocampi (ANOVA, $p < 0.0001$) compared to controls. The normalized values are also presented by the study group only, because they did not differ between women and men. After the normalization, the hippocampus still was significantly smaller bilaterally in the AD group ($p < 0.0001$). In the OC group, hippocampal volumes correlated strongly to the both coronal and sagittal intracranial areas ($p < 0.001$). Figure 5 A and B depicts z-scores of the right and left hippocampus normalized to coronal intracranial area for the study groups.

There was no significant difference in hippocampal volumes between the older and younger controls. In absolute numbers (Table 2), virtually no difference at all was found. In the nondemented groups, age was not significantly related with the right ($r = -0.19$) or left ($r = -0.19$) hippocampal volume. Because the hippocampus did not differ significantly in size between the OC and AAMI groups, we combined these non-demented groups into a large control group for further analyses. Yet, in order to get the groups age-matched, the young controls were excluded from the discrimination analyses. The combined OC–AAMI group thus formed consists of 86 study subjects. After the combination the gender difference, due to the female overrepresentation in the AAMI group, was no longer significant.

The value of volume measurements to differentiate AD patients from controls was tested in five discriminant function analyses (Table 3). The analyses included: 1) the raw volumes of the hippocampus; 2) the raw volumes and gender; 3) hippocampal

TABLE 2

RAW AND NORMALIZED HIPPOCAMPAL VOLUMES AND HIPPOCAMPAL AREAS (mm^2). RAW VOLUMES (mm^3) ARE PRESENTED BY GROUP AND GENDER

	AD	AAMI	OC	YC	ANOVA
HR, all	2322 ± 683†	3319 ± 447	3394 ± 519	3554 ± 607	41.5*
HR, male	2488 ± 717†	3613 ± 211	3459 ± 551	3700 ± 617	17.4*
HR, female	2150 ± 612†	3218 ± 464	3340 ± 497	3408 ± 592	29.5*
HL, all	2054 ± 553†	3100 ± 399	3189 ± 536	3298 ± 467	59.8*
HL, male	2140 ± 561†	3360 ± 306	3435 ± 536	3478 ± 466	34.5*
HL, female	1965 ± 542†	3011 ± 392	2985 ± 453	3118 ± 413	33.8*
HR/ICA1	17.9 ± 4.7†	26.7 ± 3.1	26.8 ± 3.7	27.6 ± 3.9	60.1*
HL/ICA1	15.9 ± 3.9†	25.0 ± 3.0	25.1 ± 3.6	25.6 ± 2.8	84.8*
HR/ICA2	11.7 ± 3.3†	17.7 ± 2.8	17.3 ± 2.5	17.3 ± 2.9	45.7*
HL/ICA2	10.4 ± 2.7†	16.6 ± 2.5	16.2 ± 2.2	16.1 ± 2.3	68.3*
HRarea	99 ± 35‡	112 ± 43‡	140 ± 32	142 ± 39	11.3*
HLarea	79 ± 30‡	101 ± 31‡	121 ± 37	129 ± 45	15.3*

Results are expressed as mean ± SD. HR right and HL left hippocampal volume; ICA1 and ICA2 coronal and sagittal intracranial area. ANOVA over the study groups, *$p < 0.0001$; Duncan post hoc analysis †$p < 0.0001$ AD vs. all other groups; ‡$p < 0.05$ AD vs. OC and YC.

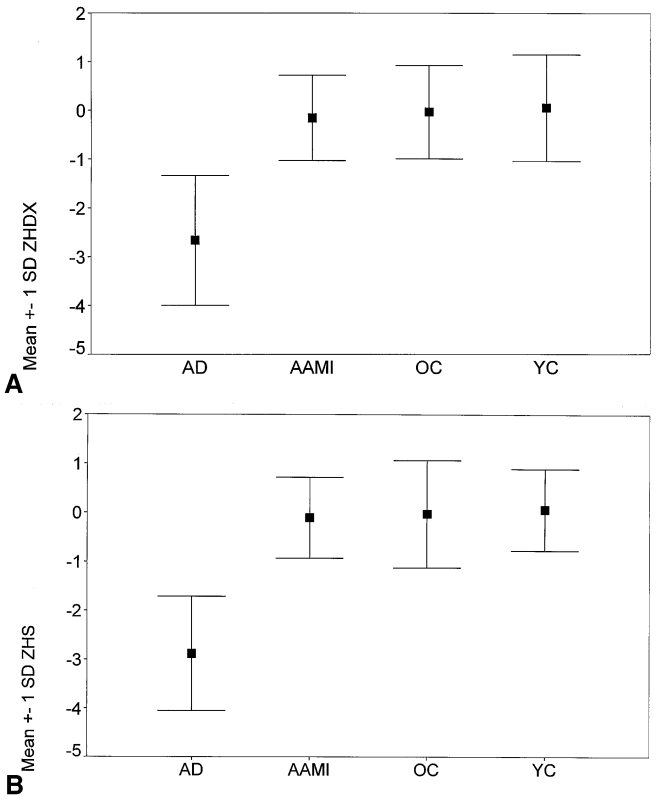

FIG. 5. *A* and *B*. Mean ± 1 standard deviation (SD) of Z-scores for the right (ZHDX) and left (ZHS) hippocampus normalized for coronal intracranial area in patients with Alzheimer's disease (AD), age-associated memory impairment (AAMI), old controls (OC), and young controls (YC).

volumes normalized for the coronal ICA1; 4) volumes normalized for the sagittal ICA2, and 5) emphasized volumes normalized to brain area/ICA1.

The analysis including the volumes of both the right and left hippocampus yielded a sensitivity of 83.6% and a specificity of 89.4% (χ^2 104.8, df 2, Wilks' lambda 0.46, $p < 0.0001$). The analysis explained 54% of the variance between groups and the volume of the left hippocampus alone explained 53%. The analyses including normalized volumes resulted in similar high sensitivity and specificity. The highest sensitivity (94.4%) was achieved in the analysis in which the hippocampal volumes were normalized taking into account the intracranial area and brain area.

We also tested whether the discriminant function works in another sample of AD patients and controls. For that purpose, a random 50% sample was drawn from the study population to generate a discriminant function and this function was then tested in the other half of the study subjects. In this analysis, we included right and left hippocampal volume normalized for coronal intracranial area and used Wilks' method. The discriminant function generated in the first half of the subjects yielded sensitivity of 81%, specificity of 90%, and overall correct classification 86.4%. When the discriminant function was used in the second half of the study population, a sensitivity of 82.1%, specificity of 97.7%, and overall correct classification of 91.6 was achieved.

Discrimination Based on Hippocampal Area

The hippocampal areas were significantly smaller on the right and left for AD patients and AAMI subjects than older and younger controls (ANOVA/Duncan $p < 0.05$). On the left, AD patients also differed significantly from AAMI subjects (Duncan $p < 0.05$). The areas of the hippocampus measured at the level of the anterior commissure on both sides correlated to the actual volume on the right ($r = 0.70$, $N = 125$, $p < 0.0001$) and on the left ($r = 0.71$). The discriminant function analysis using unnormalized areas between the AD and the OC group, without AAMI, produced a sensitivity of 76.9% and specificity of 72.4% resulting to an overall correct classification of only 75.3% (χ^2 36.1, df 2, Wilks' lambda 0.63, $p < 0.0001$). After normalization, the overall accuracy was improved to 80.0%.

Analysis of the Incorrectly Classified

The main results of the analysis of the incorrectly classified are presented in Table 4. The AD patients who were correctly and incorrectly classified in the best discriminant function analysis did not differ significantly in age, sex, age of onset, duration of the disease, education, extrapyramidal signs assessed by the Webster scale, depressive symptoms evaluated by the Hamilton scale, ischemic scores, or frequency of a positive family history. As expected the hippocampal volumes were significantly larger in the ADMC than in the ADCC group ($p < 0.0001$). The ADMC patients also had significantly higher MMSE scores ($p < 0.05$) and less severe memory impairment in tests assessing delayed recall of the Heaton Visual Reproduction test ($p < 0.05$) and the story ($p = 0.0002$) than ADCC patients. Otherwise, the profile of cognitive deficits were similar. The evaluation of follow-up data of the nine ADMC patients showed that only two patients displayed typical features of AD. Of the other seven patients, one patient had

TABLE 3

SENSITIVITY AND SPECIFICITY OF HIPPOCAMPAL VOLUMETRIC MEASURES IN DISCRIMINANT FUNCTION ANALYSES BETWEEN ALZHEIMER PATIENTS AND NONDEMENTED ELDERLY SUBJECTS

Measures	Sensitivity	Specificity	Overall correct classification
Raw volumes	46/55 (83.6)	76/85 (89.4)	122/140 (87.1)
Raw volumes and gender	49/55 (89.1)	79/85 (92.9)	128/140 (91.4)
Volumes/ICA1	46/55 (83.6)	77/83 (92.9)	123/138 (89.1)
Volumes/ICA2	46/55 (83.6)	78/84 (92.8)	124/139 (89.2)
Volumes/brain area/ICA1	51/55 (94.4)	75/83 (90.4)	126/138 (92.0)

Results are presented as number of subjects, percentage in parentheses.

TABLE 4
ANALYSIS OF THE INCORRECTLY CLASSIFIED

	Alzheimer patients Correctly classified	Alzheimer patients Misclassified	Controls Correctly classified	Controls Misclassified
Age, years	70 ± 8	70 ± 9	70 ± 5	75 ± 3*
Mini-Mental Status	21.4 ± 3.7	24.1 ± 2.6*	28.0 ± 1.5	28.8 ± 0.4
BCRS	23.8 ± 6.7	21.6 ± 6.3	NA	NA
Webster	2.1 ± 2.7	1.9 ± 2.7	NA	NA
Hamilton	3.5 ± 3.4	2.4 ± 2.2	0.3 ± 0.6	0.6 ± 0.5
Right hippocampus	2164 ± 622	3131 ± 244***	3431 ± 434	2576 ± 293***
Left hippocampus	1880 ± 398	2942 ± 341***	3214 ± 435	2385 ± 142***
Vocabulary	15.4 ± 6.4	15.6 ± 4.0	45 ± 8	45 ± 10
List learning	17.7 ± 8.4	20.7 ± 8.7	NA	NA
Buschke	NA	NA	39.6 ± 7.8	26.3 ± 7.7**
Story recall	3.6 ± 4.9	10.7 ± 4.6**	NA	NA
Heaton	0.8 ± 1.6	2.0 ± 1.9*	9.2 ± 3.9	8.0 ± 6.3
Verbal fluency	20.8 ± 13.5	20.2 ± 12.1	43.3 ± 14.2	52.6 ± 26.6
Trail making A	110.0 ± 38.6	122.2 ± 36.4	50 ± 19	35 ± 8
Trail making B	276.8 ± 58.0	287.8 ± 36.6	113 ± 61	98 ± 26
Familial dementia	39%	56%	–	–

Results are presented as means ± standard deviation. NA, not available. The control group includes old control and AAMI subjects. ANOVA correctly classified vs. misclassified within a diagnostic category; *$p < 0.05$, **$p < 0.01$, ***$p < 0.0001$.

experienced a vascular insult after the examination, one patient had strong extrapyramidal signs and may represent Lewy body disease, two patients had a strong family history for AD but a very mild possible AD, scoring 26 and 27 points in the MMSE examination, and no deterioration during the follow-up. Three patients exhibited frontal features. The AD patients of this study have been in the follow-up in our clinic and detailed investigations including MRI have been repeated 3 years later; these 3-year data are currently being analyzed. No follow-up data for the controls or AAMI subjects was available at the time of the analysis.

The incorrectly classified controls were significantly older than those correctly classified. They also had smaller hippocampal volumes ($p < 0.0001$), more severe memory decline evident from the Buschke Selective reminding test total ($p < 0.01$) and long-term scores ($p < 0.001$). Performance in other cognitive domains was equal.

DISCUSSION

Sensitivity and Specificity

In this study, we focused on MRI of the hippocampus and its use in the diagnosis of early AD. Hippocampus was chosen as a region of interest because it has a pivotal role in certain type of memory functioning (50), and because it is among the primary areas affected by AD (5,24). The hippocampus was atrophied in the AD patients, but not in AAMI subjects or the controls. In the group of 105 nondemented subjects, there was no significant correlation between hippocampal volume and age. The discriminative value of volumetric hippocampal atrophy in this study, the overall accuracy of 92.0%, shows that the volumetry provides additional objective data to support the clinical diagnosis of AD. The power of the hippocampal volumetry to discriminate AD patients from controls was further supported by testing a discriminant function generated in a random 50% sample of the study population in another half of the patients and controls. This analysis showed overall correct classification of 91.6% of patients and controls of the second half. Compared to volumetry, the measurement of hippocampal area was less accurate, producing a correct classification of only 80%.

Methodological Considerations

In this study, 160 well-documented subjects were examined. The number is by far the largest in any volumetric MRI study of AD or AAMI. The AD patients represented early stages of AD, having been recently diagnosed. The diagnosis was based on the NINCDS-ADRDA criteria, and the AAMI patients met the strict NIMH criteria for AAMI. The AAMI patients and the old controls were drawn from a population based study cohort. All the elderly study groups received not only careful clinical neurological examination and MRI, but also an additional extensive neuropsychological test battery, as earlier described in the materials and methods. The young controls were interviewed healthy members of staff, consisting mainly of doctors and medical students. The elderly study groups were well matched, with two exceptions. First, in the AAMI group, women were overrepresented. Yet, because in the combined control group this difference vanished, and so did the differences due to gender after normalization, we fail to see a source of bias here. Second, the AD group had received less education than the other groups. In general, in the age group that our AD patients, AAMI subjects, and elderly controls represent, the level of education is clearly lower in Finland than, for example, in the USA, the basic elementary education in this age group being 6 years. There is no specific explanation, e.g., based on selection of cases, why our AD and AAMI subjects tended to have lower level of education than the OC group.

To obtain accurate and objective MRI data, the entire volume of the hippocampus was measured using thin, contiguous slices. A recent study showed that the use of slice thickness varying between 1–5 mm has no effect on the volume on MR images, but the thinner the slices, the less the volume is affected by possible false estimates (36). In this study, the measurements were performed blinded by a single experienced rater. In general, use of T_1-weighted images oriented perpendicular to the hippocampus,

has made measurement of the entire hippocampus, including the anterior part, not only possible, but also reliable (1,6,27,35).

We experimented with two different methods of normalization to count for all three dimensions of the head to avoid systemic biases due to simple methodology. In the third method, the effect of general brain atrophy was evaluated by multiplying hippocampal volumes by brain area/ICA1 relationship. This yielded the best sensitivity in the study. In general, normalization methods used in this study did not notably improve the accuracy. This is not completely unexpected, much because the rate of hippocampal volume loss in AD is as much as about 1/3 of the volume of healthy subjects. The significance of proper normalization must, however, not be underestimated, but be considered as an important procedure due to strong correlation between the hippocampus and the head size. The sagittal ICA (ICA 2) has been previously validated as a proper method for normalization (16), and judging by this study, the coronal ICA (ICA 1) appears to be similarly valid a method as well. The discriminative accuracy was just as powerful, and correlation to hippocampal volumes were similar with both ICAs. Instead, normalization to brain area is less reliable for normalization of atrophic changes when atrophy presents as reduction in volume, because it itself is vulnerable to both physiological and neurodegenerative atrophy. Therefore, an index between two atrophying structures might end up with worse accuracy. In contrast, normalization by brain area might improve the accuracy of atrophic changes presenting as dilatation, or to provide information in comparison to general atrophy.

In previous literature, the used normalization variables have varied remarkably. An often used but seldom described normalization variable is the intracranial volume. The method, if merely stated, is not reproducible, because the volume can be delineated in several ways, and because delineation of a bony structure to some extent is arbitrary due to improper visualization of compact bone on MR images. In this study, the intracranial area is drawn along the most clear signal caused by subcutaneous (s.c.) adipose tissue. Therefore, appropriate name for these methods, including calvarium, might be the cranial area. In order to obtain a widely accepted imaging protocol to define normal and pathological ranges for the hippocampal volumes and for pooling comparable data from multicenter studies, a commonly accepted imaging protocol, including normalization, is needed.

Previously, the imaging studies of AD have produced most varying results. In 1992, the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), declared standardized imaging and reporting of the MRI findings to be unsatisfactory (10). The use of hippocampal imaging has similarly produced most discrepant results. However, a critical evaluation of this material reveals an obvious, yet an important pattern: the studies that have measured the entire volume of the hippocampus (27,29,30,35,37), have, without exceptions, resulted in good discriminative accuracy between AD patients and the controls. In contrast, results from studies using simple or subjective measurements have varied from accuracy of 100% (48,56) to that less than 50% (13,34). Some studies have settled somewhere in the middle (9,17,41). A recent study of various linear measurements was recently published by Frisoni et al. (17), in which the sensitivity of various measurements ranged from 24–74%. In essence, to obtain reliable data, the entire volume of hippocampus must be measured, leaving but little room for bias or subjective judgments.

Volume of the Hippocampus and Aging

In this study, and in this age span (21–79), the age had no effect on the hippocampal volumes of the nondemented subjects. This in spite of the fact that the younger control group was highly educated, "super normal" control group. As opposite results have been published, several of those may be explained similarly to those presented above: methodological issues account for the observed result. The result, in fact, is also very much in concordance with recent data suggesting that medial temporal lobe cortical areas are only minimally affected by normal aging. First, there are several previous volumetric MRI studies that have reported that the size of the hippocampus may remain unaffected by normal aging (2,11,26,51). The finding is further supported by stereological histopathological studies that have reported only minimal neuronal loss in the hippocampus and adjacent cortical areas (18,42,54,55). Finally, memory functions considered mediated by the hippocampus—encoding, consolidation and short-term storage of newly acquired data (50), the function of which may be evaluated by tests assessing delayed recall—display an identical trend. The tests assessing delayed recall have been documented to be spared by aging, and to be sensitive indicators of dementia (21,22,43,53). Delayed recall performance has also been shown to correlate with hippocampal volumes (12,35).

The AAMI Issue

As the AAMI issue remains controversial (25), the important contribution of the present study is that persons who have benign subjective and objective memory impairment, AAMI, can be reliably separated from those having early AD by hippocampal volumetry. This sparing of the volume of the hippocampus supports the assumption the that concept of AAMI is related rather to normal aging than to dementia. The follow-up data of the AAMI subjects in this study was not available yet. However, in our previous study, less than 10% of subjects classified as having AAMI became demented in a 3.6-year follow-up (25). Also, in a previous study, the volume of hippocampus in AAMI was spared, but the normal right-left asymmetry was diminished (49). In another study, the volumes of the hippocampi of AAMI subjects settled between the volumes of AD patients and the controls (40). The background of this discrepancy is unclear, but the number of AAMI subjects in this study exceeded the number of the subjects in the study of Parnetti and colleagues. It may be also emphasized that the AAMI subjects in this study were population drawn, i.e., not subjects seeking consultation from the memory clinic.

Analysis of the Incorrectly Classified

Analysis of the incorrectly classified was performed to further consider the ultimate discriminative accuracy of the volumetry. Despite the careful screening described above, we admit that this study, or any study this size, is likely to be contaminated with some false positive or negative cases, particularly when dealing with early AD. The NINCDS-ADRDA criteria are not perfect, but provide diagnostic accuracy of 80–90% (3,33). In our institute, the neuropathologically confirmed accuracy of NINCDS-ADRDA criteria to diagnose probable AD has been reported to be 96% (32). According to CDR score, the majority of our AD patients in this study had mild, some even questionable, dementia. Nine controls were incorrectly classified by MRI volumetry. Considering that they were older, their performance in Buschke Selective Reminding test was worse, the prevalence of AD increasing with age, reaching 13% at the age of 80 (28), and the hippocampal atrophy may precede the symptoms of dementia (15), these subjects may possibly represent preclinical dementia. Also, nine patients primarily classified having AD were incorrectly classified. They performed significantly better in the MMSE and tests assessing delayed recall, but their deficits in other cognitive domains were similar to those of ADCC. Individual reevaluation showed that seven out of nine AD patients had atypical features or very mild

disease with nonprogressive clinical course. Thus, it would be tempting to assume that the true discriminative accuracy may be higher than reported.

In conclusion, volumetric atrophy of the hippocampus is a highly sensitive indicator of AD early in the course of the disease and provides additional data to support the clinical diagnosis. Both elderly controls and AAMI subjects were classified with over 90% accuracy. Instead, the accuracy by hippocampal area measurement was only 80%. It appears that simple or subjective measurements of the hippocampus are subject to various sources of bias and do not provide the same accuracy the volumetry does.

ACKNOWLEDGEMENTS

This study was supported by the Medical Research Council of the Academy of Finland, the Finnish Medicine Foundation, the North Savo Fund of the Finnish Cultural Foundation, and the Finnish Neurology Foundation.

REFERENCES

1. Bartzokis, G.; Mintz, J.; Marx, P.; Osborn, D.; Gutkind, D.; Chiang, F.; Phelan, CK.; Marder, S. R. Reliability of in vivo measures of hippocampus and other brain structures using MRI. Magn. Reson. Imaging 11:993–1006; 1996.
2. Bhatia, S.; Bookeheimer, S. Y.; Gaillard, W. D.; Theodore, W. H. Measurement of whole temporal lobe and hippocampus for MR volumetry: normative data. Neurology 43:2006–2010; 1993.
3. Blacker, D.; Albert, M. S.; Basset, S. S.; Go, R. C. P.; Harrell, L. E.; Folstein, M. F. Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease. The National Institute of Mental Health Genetics Initiative. Arch. Neurol. 51:1198–1204; 1994.
4. Borkowski, J. G.; Benton, A. L.; Spreen, O. Word fluency and brain damage. Neuropsychologia 5:135–140; 1967.
5. Braak, H.; Braak, E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 82:239–259; 1991.
6. Bronen, R. A.; Cheung, G. MRI of the temporal lobe: normal variations, with special reference toward epilepsy. Magn. Reson. Imaging 9:501–507; 1991.
7. Buschke, H.; Fuld, P. Evaluating storage, retention and retrieval in disordered memory and learning. Neurology 24:1019–1025; 1974.
8. Crook, T. H.; Bartus, R. T.; Ferris, S. H.; Whitehouse, P.; Cohen, G. D.; Gershon, S. Age-associated memory impairment: Proposed diagnostic criteria and measures of clinical change - Report of a National Institute of Mental Health work group. Dev. Neuropsychol. 2:261–276; 1986.
9. Cuénod, C.-A.; Denys, A.; Michot, J.-L.; Jehenson, P.; Forette, F.; Kaplan, D.; Syrota, A.; Boller, F. Amygdala atrophy in Alzheimer's disease. Arch. Neurol. 50:941–945; 1993.
10. Davis, P. C.; Gray, L.; Albert, M.; Wilkinson, W.; Hughes, J.; Heyman, A.; Gado, M.; Kumar, A. J.; Destian, S.; Lee, C.; Duvall, E.; Kido, D.; Nelson, M. J.; Bello, J.; Weathers, S.; Jolesz, F.; Kikinis, R.; Brooks, M. The Concortium to Establish a Registry for Alzheimer's Disease (CERAD). Part III. Reliability of standardized MRI evaluation of Alzheimer's disease. Neurology 42:1676–1680; 1992.
11. DeCarli, C.; Murphy, D. G. M.; Gillette, J. A.; Haxby, J. V.; Teichberg, D.; Schapiro, M. B.; Horwitz, B. Lack of age-related differences in temporal lobe volume of very healthy adults. Am. J. Neuroradiol. 15:689–696; 1994.
12. Deweer, B.; Lehéricy, S.; Pillon, B.; Baulac, M.; Chiras, J.; Marsault, C.; Agid, Y.; Dubois, B. Memory disorders in probable Alzheimer's disease: The role of hippocampal atrophy as shown with MRI. J. Neurol. Neurosurg. Psychiatry 58:590–597; 1995.
13. Erkinjuntti, T.; Lee, D. H.; Gao, F.; Steenhuis, R.; Eliasziw, M.; Fry, R.; Merskey, H.; Hachinski, V. Temporal lobe atrophy on magnetic resonance imaging in the diagnosis of early Alzheimer's disease. Arch. Neurol. 50:305–310; 1993.
14. Folstein, M. F.; Folstein, S. E.; McHugh, P. R. "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. J. Psychiatry Res. 12:189–198; 1975.
15. Fox, N. C.; Warrington, E. K.; Freeborough, P. A.; Hartikainen, P.; Kennedy, A. M.; Stevens, J. M.; Rossor, M. N. Presymptomatic hippocampal atrophy in Alzheimer's disease. A longitudinal MRI study. Brain 119:2001–2007; 1996.
16. Free, S. L.; Bergin, P. S.; Fish, D. R.; Cook, M. J.; Shorvon, S. D.; Stevens, J. M. Methods for normalization of hippocampal volumes measured with MR. Am. J. Neuroradiol. 16:637–643; 1995.
17. Frisoni, G. B.; Beltramello, A.; Weiss, C.; Geroldi, C.; Bianchetti, A.; Trabucchi, M. Linear measures of atrophy in mild Alzheimer's disease. Am. J. Neuroradiol. 17:913–923; 1996.
18. Gómez-Isla, T.; Price, J. L.; McKeel, D. W.; Morris, J. C.; Growdon, J. H.; Hyman, B. T. Profound loss of layer II entrohinal cortex neurons occurs in very mild Alzheimer's disease. J. Neurosci. 16:4491–4500; 1996.
19. Hamilton, M. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23:56–62; 1960.
20. Helkala, E.-L.; Laulumaa, V.; Soininen, H.; Riekkinen, P. J. Recall and recognition on memory in patients with Alzheimer's and Parkinson's diseases. Ann. Neurol. 24:214–217; 1988.
21. Hodges, J. R.; Patterson, K. Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. Neuropsychologia 33:441–459; 1995.
22. Howieson, D. B.; Holma, L. A.; Kaye, J. A.; Oken, B. S.; Howieson, J. Neurologic function in the optimally healthy oldest old: Neuropsychological evaluation. Neurology 43:1882–1886; 1993.
23. Hughes, C.; Berg, L.; Danziger, L.; Cohen, L. A.; Martin, R. L. A new clinical rating scale for staging dementia. Br. J. Psychiatry 140:566–572; 1982.
24. Hyman, B. T.; Damasio, A. R.; Van Hoesen, G. W.; Barnes, C. L. Alzheimer's disease: Cell-specific pathology isolates the hippocampal formation. Science 225:1168–1170; 1984.
25. Hänninen, T.; Hallikainen, M.; Koivisto, K.; Helkala, E.-L.; Reinikainen, K. J.; Soininen, H.; Mykkänen, L.; Laakso, M.; Pyörälä, K.; Riekkinen, P. J. Sr. A follow-up study of age-associated memory impairment: Neuropsychological predictors of dementia. J. Am. Geriatric Soc. 43:1007–1015; 1995.
26. Jack, C. R. Jr.; Twomey, C. K.; Zinsmeister, A. R.; Sharbrough, F. W.; Petersen, R. C.; Cascino, G. D. Anterior temporal lobes and hippocampal formations: Normative volumetric measurements from MR images in young adults. Radiology 172:549–554; 1989.
27. Jack, C. R. Jr.; Petersen, R. C.; O'Brien, P. C.; Tangalos, E. G. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology 42:183–188; 1992.
28. Juva, K.; Sulkava, R.; Erkinjuntti, T.; Valvanne, J.; Tilvis, R. Prevalence of dementia in the city of Helsinki. Acta Neurol. Scand. 87:106–110; 1993.
29. Kesslak, J. P.; Nalcioglu, O.; Cotman, C. W. Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease. Neurology 41:51–54; 1991.
30. Killiany, R. J.; Moss, M. B.; Albert, M. S.; Sandor, T.; Tieman, J.; Jolesz, F. Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. Arch. Neurol. 50: 949–954; 1993.
31. Koivisto, K.; Reinikainen, K. J.; Hänninen, T.; Vanhanen, M.; Helkala, E.-L.; Mykkänen, L.; Laakso, M.; Pyörälä, K, Riekkinen, P. J. Sr. Prevalence of age-associated memory impairment in a randomly selected population from Eastern Finland. Neurology 45:741–747; 1995.
32. Kosunen, O.; Soininen, H.; Paljärvi, L.; Heinonen, O.; Talasniemi, S.; Riekkinen, P. J. Sr. Diagnostic accuracy of Alzheimer's disease. A neuropathological study. Acta Neuropathol. 91:185–193; 1996.
33. Kukull, W. A.; Larson, E. B.; Reifler, B. V.; Lampe, T. H.; Yerby, M. S.; Hughes, J. P. The validity of 3 clinical diagnostic criteria for Alzheimer's disease. Neurology 40:1364–1369; 1990.
34. Laakso, M. P.; Soininen, H.; Partanen, K.; Hallikainen, M.; Lehtovirta,

M.; Hänninen, T.; Vainio, P.; Riekkinen, P. J. Sr. The interuncal distance in Alzheimer's disease and age-associated memory impairment. Am. J. Neuroradiol. 16:727–734; 1995.
35. Laakso, M. P.; Soininen, H.; Partanen, K.; Helkala, E.-L.; Hartikainen, P.; Vainio, P.; Riekkinen, P. J. Sr. Volumes of hippocampus, amygdala and frontal lobes in the MRI-based diagnosis of early Alzheimer's disease: Correlation with memory functions. J. Neural Transm. [P-D Sect]. 9:73–86; 1995.
36. Laakso, M. P.; Juottonen, K.; Partanen, K.; Vainio, P.; Soininen, H. MRI volumetry of the hippocampus: The effect of slice thickness on volume formation. Magn. Reson. Imaging 15:263–265; 1997.
37. Lehéricy, S.; Baulac, M.; Chiras, J.; Piérot, L.; Marin, N.; Pillon, B.; Deweer, B.; Dubois, B.; Marsault, C. Amygdalohippocampal MR volume measurements in the early stages of Alzheimer disease. Am. J. Neuroradiol. 15:927–937; 1994.
38. McKhann, G.; Drachman, D.; Folstein, M.; Katzman, R.; Price, D.; Stadlan, E. M. Clinical diagnosis of Alzheimer's disease: Report of NINCDS/ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. Neurology 34:939–944; 1984.
39. Millber, W. P.; Hebben, N.; Kaplan, E. The Boston process approach to neuropsychological assessment. In: Grant, I; Adams, K.M., eds., Neuropsychological Assessment of Neuropsychiatric Disorders. New York: Oxford University Press, 1986:65–86.
40. Parnetti, L.; Lowenthal, D. T.; Presciutti, O.; Pelliccioli, G.; Palumbo, R.; Gobbi, G.; Chiarini, P.; Palumbo, B.; Tarducci, R.; Senin, U. ^{1}H-MRS, MRI-based hippocampal volumetry, and ^{99m}Tc-HMPAO-SPECT in normal aging, age-associated memory impairment, and probable Alzheimer's disease. J. Am. Geriatric Soc. 44:133–138; 1996.
41. Pearlson, G. D.; Gordon, J. H.; Powers, R. E.; Barta, P. E.; Camargo, E. E.; Chase, G. A.; Noga, J. T.; Tune, L. E. Quantitative changes in mesial temporal volume, regional cerebral blood flow, and cognition in Alzheimer's disease. Arch. Gen. Psychiatry 49:402–408; 1992.
42. Peters, A.; Rosene, D. L.; Moss, M. B.; Kemper, T. L.; Abraham, C. R.; Tigges, J.; Albert, M. S. Neurobiological bases of age-related cognitive decline in the Rhesus monkey. J. Neuropathol. Exp. Neurol. 55:861–874; 1996.
43. Petersen, R. C.; Smith, G.; Kokmen, E.; Ivnik, R. J.; Tangalos, E. G. Memory function in normal aging. Neurology 42:396–401; 1992.
44. Reisberg, B.; Schenck, M. K.; Ferris, S. H. The brief cognitive rating scale (BCRS): Findings in primary degenerative dementia (PDD). Psychopharmacol. Bull. 19:734–739; 1983.
45. Reitan, R. M. Validity of the Trail Making test as an indicator of organic brain damage. Perc. Mot. Skills 8:271–276; 1958.
46. Rosen, W. G.; Terry, R. D.; Fuld, P. A.; Katzman, R.; Beck, A. Pathological verification of ischemic score in differentiation of dementias. Ann. Neurol. 17:486–488; 1980.
47. Russell, E. W. A multiple scoring method for the assessment of complex memory functions. J. Cons. Clin. Psychology 43:800–809; 1975.
48. Seab, J. P.; Jagust, W. J.; Wong, S. T. S.; Roos, M. S.; Reed, B. R.; Budinger, T. F. Quantitative NMR measurements of hippocampal atrophy in Alzheimer's disease. Magn. Reson. Med. 8:200–208; 1988.
49. Soininen, H. S.; Partanen, K.; Pitkänen, A.; Vainio, P.; Hänninen, T.; Hallikainen, M.; Koivisto, K.; Riekkinen, P. J. Sr. Volumetric MRI analysis of the amygdala and the hippocampus in subjects with age-associated memory impairment: Correlation to visual and verbal memory. Neurology 44:1660–1668; 1994.
50. Squire, L. R.; Zola-Morgan, S. The medial temporal lobe memory system. Science 253:1380–1386; 1991.
51. Sullivan, E. V.; Marsh, L.; Mathalon, D. H.; Lim, K. O.; Pfefferbaum, A. Age-related decline in MRI volumes of temporal lobe gray matter but not hippocampus. Neurobiol. Aging 16:591–606; 1995.
52. Webster, D. D. Clinical analysis of the disability on Parkinson's disease. Mod. Treat. 5:257–262; 1968.
53. Welsh, K. A.; Butters, N.; Hughes, J. P.; Mohs, R. C.; Heyman, A. Detection and staging of dementia in Alzheimer's disease: Use of neuropsychological measures developed for the Consortium to establish a Registry for Alzheimer's Disease. Arch. Neurol. 49:448–452; 1992.
54. West, M. J.; Coleman, P. D.; Flood, D. G.; Troncoso, J. C. Differences in the pattern of hippocampal neuronal loss in normal ageing and Alzheimer's disease. Lancet 344:769–772; 1994.
55. Wickelgren, I. Is hippocampal cell death a myth? Science 271:1229–1330; 1996.
56. Yamada, N.; Tanabe, H.; Kazui, H.; Ikeda, M.; Hashimoto, M.; Nakagawa, Y.; Wada, Y.; Okuda, J.-I. Longitudinal neuropsychological and quantitative MRI assessments in early Alzheimer's disease. Alzheimer's Res. 2:29–36; 1996.