

Imaging markers for Alzheimer disease

Which vs how



Giovanni B. Frisoni, MD
Martina Bocchetta, MS
Gael Chételat, PhD
Gil D. Rabinovici, MD
Mony J. de Leon, EdD
Jeffrey Kaye, MD
Eric M. Reiman, MD
Philip Scheltens, MD,
PhD
Frederik Barkhof, MD,
PhD
Sandra E. Black, MD
David J. Brooks, MD
Maria C. Carrillo, MD,
PhD
Nick C. Fox, MD, PhD
Karl Herholz, MD
Agneta Nordberg, MD,
PhD
Clifford R. Jack, Jr., MD
William J. Jagust, MD
Keith A. Johnson, MD,
PhD
Christopher C. Rowe,
MD, PhD
Reisa A. Sperling, MD
William Thies, PhD
Lars-Olof Wahlund, MD,
PhD
Michael W. Weiner, MD
Patrizio Pasqualetti, PhD
Charles DeCarli, MD
For ISTAART's

NeuroImaging
Professional Interest
Area

Correspondence to
Dr. Frisoni:
gfrisoni@fatebenefratelli.it

Supplemental data at
www.neurology.org

ABSTRACT

Revised diagnostic criteria for Alzheimer disease (AD) acknowledge a key role of imaging biomarkers for early diagnosis. Diagnostic accuracy depends on *which* marker (i.e., amyloid imaging, ^{18}F -fluorodeoxyglucose [FDG]-PET, SPECT, MRI) as well as *how* it is measured ("metric": visual, manual, semiautomated, or automated segmentation/computation). We evaluated diagnostic accuracy of marker vs metric in separating AD from healthy and prognostic accuracy to predict progression in mild cognitive impairment. The outcome measure was positive (negative) likelihood ratio, LR+ (LR−), defined as the ratio between the probability of positive (negative) test outcome in patients and the probability of positive (negative) test outcome in healthy controls. Diagnostic LR+ of markers was between 4.4 and 9.4 and LR− between 0.25 and 0.08, whereas prognostic LR+ and LR− were between 1.7 and 7.5, and 0.50 and 0.11, respectively. Within metrics, LRs varied up to 100-fold: LR+ from approximately 1 to 100; LR− from approximately 1.00 to 0.01. Markers accounted for 11% and 18% of diagnostic and prognostic variance of LR+ and 16% and 24% of LR−. Across all markers, metrics accounted for an equal or larger amount of variance than markers: 13% and 62% of diagnostic and prognostic variance of LR+, and 29% and 18% of LR−. Within markers, the largest proportion of diagnostic LR+ and LR− variability was within ^{18}F -FDG-PET and MRI metrics, respectively. Diagnostic and prognostic accuracy of imaging AD biomarkers is at least as dependent on how the biomarker is measured as on the biomarker itself. Standard operating procedures are key to biomarker use in the clinical routine and drug trials.

Neurology® 2013;81:487–500

GLOSSARY

AA = Alzheimer's Association; **AD** = Alzheimer disease; **ANOVA** = analysis of variance; **FDG** = fluorodeoxyglucose; **ISTAART** = International Society to Advance Alzheimer's Research and Treatments; **LR** = likelihood ratio; **MCI** = mild cognitive impairment; **NIA** = National Institute on Aging; **NINCDS-ADRDA** = National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association; **npMCI** = nonprogressed mild cognitive impairment; **pMCI** = progressed mild cognitive impairment; **sc-SPM** = single-case statistical parametric mapping; **SOP** = standard operating procedure.

Recent acquisitions on the pathophysiology and natural history of Alzheimer disease (AD) have led researchers to propose alternatives to the traditional NINCDS-ADRDA diagnostic criteria. The International Working Group^{1,2} and National Institute on Aging–Alzheimer's Association (NIA-AA) criteria^{3–5} assign a key pathogenetic role to cerebral β -amyloidosis and neurodegeneration,

From the LENITEM (Laboratory of Epidemiology, Neuroimaging and Telemedicine) (G.B.F., M.B.), IRCCS, S. Giovanni di Dio, Fatebenefratelli Brescia, Italy; INSERM (G.C.), U1077, Caen; Université de Caen Basse-Normandie (G.C.), UMR-S1077, Caen; Ecole Pratique des Hautes Etudes (G.C.), UMR-S1077, Caen; CHU de Caen (G.C.), U1077, Caen, France; Memory and Aging Center, Department of Neurology (G.D.R.), University of California, San Francisco (C.D., M.W.W.); New York University School of Medicine (M.J.d.L.), Center for Brain Health, New York, NY; Oregon Health & Science University (J.K.), Portland; the Portland Veterans Affairs Medical Center (J.K.); Banner Alzheimer's Institute (E.M.R.), Phoenix, AZ; Department of Neurology and Alzheimer Center (P.S.), and Departments of Radiology and Nuclear Medicine (F.B.), VU University Medical Center, Amsterdam; the Netherlands; Department of Medicine (Neurology) (S.E.B.), University of Toronto, Sunnybrook Research Institute, Toronto, Canada; Division of Brain Sciences (D.J.B.), Faculty of Medicine, Imperial College, London, UK; Aarhus University (D.J.B.), Denmark; Medical & Scientific Relations (M.C.C., W.T.), Alzheimer's Association, Chicago, IL; University College London Institute of Neurology (N.C.F.), London; Wolfson Molecular Imaging Centre (K.H.), Institute of Brain, Behaviour and Mental Health, University of Manchester, UK; Karolinska Institutet (A.N.), Karolinska University Hospital Huddinge, Stockholm, Sweden; Alzheimer Neurobiology Center (A.N.), Karolinska Institutet, Stockholm; Department of Diagnostic Radiology (C.R.J.), Mayo Clinic and Foundation, Rochester, MN; School of Public Health & Helen Wills Neuroscience Institute (W.J.J.), University of California, Berkeley; Department of Neurology (K.A.J., R.A.J.), Massachusetts General Hospital and Brigham and Women's Hospital, Boston; Department of Nuclear Medicine (C.C.R.), Centre for PET, Austin Health, Melbourne, Australia; Division of Clinical Geriatrics (L.-O.W.), NVS Department, Karolinska Institutet, Stockholm, Sweden; San Francisco Veterans Affairs Medical Center (M.W.W.); SeSMIT (Service for Medical Statistics and Information Technology) (P.P.), AfAR (Fatebenefratelli Association for Research), Fatebenefratelli Hospital, Isola Tiberina, Rome; and Unit of Clinical and Molecular Epidemiology (P.P.), IRCCS San Raffaele Pisana, Rome, Italy.

ISTAART's NeuroImaging Professional Interest Area contributors are listed on the *Neurology*® Web site at www.neurology.org.

Go to *Neurology.org* for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

hallmarked by senile plaques and neuronal tangles on microscopic examination. They further stipulate that positivity on one or more disease markers of brain amyloidosis (decreased levels of A β 42 in the CSF and increased binding of amyloid imaging agents with PET) and neuronal injury (cortical temporoparietal hypometabolism on ¹⁸F-fluorodeoxyglucose [FDG]-PET, or hypoperfusion on SPECT, medial temporal atrophy on MRI, and increased tau or phospho-tau in the CSF) is associated with high likelihood that the patient's cognitive impairment is due to AD pathology.

The view is largely shared that the criteria, although potentially applicable, are not ready to be widely used in routine clinical practice,^{6–9} although a fluorinated ligand¹⁰ is qualified by US and European Union regulatory agencies,^{11,12} and amyloid PET and hippocampal volume are qualified by the latter for enrichment in clinical trials of AD modifiers.^{13,14} None of these biomarkers, neither imaging nor fluid, is reimbursed by health care providers or third party payers. However, some specialized clinical services with the appropriate knowledge and facilities are using biomarkers as adjuncts in the diagnostic process, supporting the practical urgency of quick progression on the track of criteria validation. In this context, the intrinsic test characteristics of biomarkers will represent a key factor for successful validation.

A number of reviews are available on the diagnostic accuracy of imaging biomarkers. Reviews have generally focused on single modality markers (i.e., MRI, FDG-PET, amyloid PET, or perfusion SPECT markers), and only a few have addressed accuracy across different modalities (e.g., MRI vs FDG-PET markers). Still fewer have studied diagnostic accuracy across different operating procedures, and none has addressed diagnostic accuracy of imaging biomarkers across different modalities and operating procedures. The latter effort is important to appreciate the relevance of modality and operating procedure on diagnostic accuracy. This information will help in designing clinical research studies aimed at validating the new diagnostic criteria for AD, and contribute to the progression of imaging biomarkers from informal diagnostic adjuncts to fully validated biomarkers.

We aimed at estimating the diagnostic and prognostic accuracy of different AD imaging biomarkers (here called “markers”) and their operating procedures (here called “metrics”), and to investigate the amount and source of variance among them. This review was conceived by the Neuroimaging Professional Interest Area, a group of clinical imaging scientists borne of the Alzheimer's Imaging Consortium and the specialist branch of the International Society to Advance Alzheimer's Research and Treatment (ISTAART) of the AA, in the context of its mission to promote the appropriate use of imaging in clinical and research contexts. The views expressed herein are those of the authors and do not represent a formal position or endorsement by the AA.

METHODS Inclusion and exclusion criteria. We performed a search on the PubMed database for literature published between 1989 and April 2012, using combined specific terms of AD, accuracy, and biomarkers: “condition AND marker AND submarker AND (accuracy OR sensitivity OR specificity),” where conditions were “Alzheimer's disease” and “mild cognitive impairment,” markers were “amyloid PET,” “SPECT or SPET,” “¹⁸F-FDG PET,” “magnetic resonance,” whereas submarkers were “¹⁸F” and “¹¹C-PiB” for amyloid PET; “hippocampus,” “amygdala,” “entorhinal cortex,” and “temporal horn” for MRI; and “^{99m}Tc-HMPAO” and “^{99m}Tc-ECD or ¹²³I-IMP” for SPECT. The “related articles” feature in PubMed for the selected research studies and references of retrieved articles were also screened to maximize the probability of finding additional relevant studies. We extracted single studies from reviews and meta-analyses^{15–26} and addressed them individually. The search was limited to articles involving human subjects and written in English.

We included studies reporting sensitivity and specificity for each single analytic method for each biomarker (“metric”), and the number and the diagnosis of subjects for each comparison group. The clinical diagnosis was the comparator between different studies. For mild cognitive impairment (MCI), we included only studies that considered sensitivity as the correct classification of patients with MCI who subsequently progressed to AD dementia (pMCI) vs patients with MCI who did not progress (npMCI).

We excluded studies if they did not i) study patients with AD or MCI; ii) report numerical data for sensitivity and specificity; iii) explicitly state procedures for marker measurement; iv) assess the diagnostic performance of individual imaging biomarkers (e.g., accuracy of clinical diagnosis plus biomarkers, or a panel of biomarkers); v) disaggregate pMCI from npMCI; or vi) provide information on group size. Studies of AD vs other types of dementia were not considered because of i) the low number of available studies, and ii) the fact that we should have further disaggregated studies not only by markers by metrics but also by non-AD conditions, thus resulting in an unacceptably small group size per cell. We excluded studies comparing healthy elderly people and patients with MCI because of the huge etiologic heterogeneity of the MCI group, and studies of patients with MCI who progressed to non-AD dementias.

Metrics. The selected studies were classified based on the specific marker acquisition and analytic approach (metric) (figure 1). Metrics are described below for each marker.

Amyloid imaging agents with PET. Metrics include: i) visual read, the qualitative assessment of cortical ligand uptake for each image; ii) standardized uptake value ratio, the quantitative analysis of the ratio of cortical ligand uptake to a reference region for each image; and iii) distribution volume ratio, the quantitative analysis of the ratio of cortical ligand distribution volume to the cerebellar uptake for each image.

Temporoparietal hypometabolism on ^{18}F -FDG-PET. Metrics include: i) computer-aided visual read (Neurostat/3D-SSP, <http://www.rad.washington.edu/research/Research/groups/nbl/neurostat-3d-ssp>), which uses the 3-dimensional stereotactic surface projection technique through the Neurostat automated image analysis procedure, comparing each image on a pixel-by-pixel basis with a normative reference database, and producing parametric z score images; ii) t-sum/hypometabolic convergence index, the automated summary measures of AD-related hypometabolism based on the comparison of individual images with a normative reference dataset in a predefined AD mask (t-sum score is computed as voxel-by-voxel sum of t scores in a predefined AD-pattern mask,²⁷ whereas the hypometabolic convergence index is calculated as the inner product of the individual Z-map and a predefined AD Z-map²⁸); iii) computer-aided visual read using single-case statistical parametric mapping (sc-SPM) (<http://www.fil.ion.ucl.ac.uk/spm>), computing a score as the average metabolism on a set of meta-analytically derived regions of interest reflecting the AD hypometabolism pattern; and iv) visual read, the qualitative assessment of cortical metabolism for each image.

Temporoparietal hypoperfusion on SPECT or SPET. Metrics include: i) visual read, the qualitative assessment of cortical perfusion for each individual image; and ii) quantitative/semiquantitative assessment, the quantification of cortical perfusion for each image.

Medial temporal atrophy on structural MRI. Metrics include: i) visual read, the qualitative assessment of structure atrophy using Likert scales; ii) manual segmentation, the volumetric measurement through manual segmentation; iii) automated volumetry measurement computed through automated segmentation algorithms (FreeSurfer, which implements the subcortical segmentation by probabilistic segmentation based on a prior anatomical model^{29,30}; AdaBoost-ACM, a “machine learning” method that learns features to guide segmentation³¹; BrainVISA SASHA, the deformation constraint approach based on prior knowledge of anatomical features automatically retrieved from MRI data³²); and iv) linear measure, the manual measurement of the medial temporal lobe and the temporal horn of the lateral ventricle.

Table 1 suggests that metrics are remarkably heterogeneous for acquisition procedures, automation, stability, intensivity (in terms of human or machine time), availability of a normative population and threshold, and cost.

Outcome measure. To investigate the variability attributable to markers, submarkers, and metrics, we chose the likelihood ratio (LR). We preferred this to the more traditional sensitivity and specificity because it combines information of both sensitivity and specificity and is not affected by arbitrary thresholds that authors may choose to maximize the specificity or sensitivity of a test. Positive and negative LR (LR+ and LR−) were computed as follows: $\text{LR+} = \text{sensitivity}/(100 - \text{specificity})$ and $\text{LR-} =$

Figure 1 Markers, submarkers, and metrics reviewed in the current study

Marker	Increased binding of amyloid imaging ligands with PET		Temporoparietal hypometabolism on 18F-FDG PET	Temporoparietal hypoperfusion on SPECT		Medial temporal atrophy on MRI			
Submarker	11C-PIB	18F ligands	---	99m Tc-ECD and 123I-IMP	99m Tc-HMPAO	Amygdala	Hippocampus	Entorhinal cortex	Temporal horn
Metric	SUVR (Standardized Uptake Value Ratio)		Computer-aided visual read (Neurostat/3D-SSP)	Visual read		---		Linear measure	
	Visual read	---	Automated TP hypometabolism (t-sum/HCI)	Quantitative/semiquantitative assessment		Visual read		---	
	DVR (Distribution Volume Ratio)		Computer-aided visual read (sc-SPM)			Manual volumetry			
			Visual read			---	Automated volumetry		

3D-SSP = 3-dimensional stereotactic surface projection; ^{11}C -PiB = ^{11}C -labeled Pittsburgh compound B; $^{99\text{m}}\text{Tc}$ -ECD = technetium-99m bicitate; $^{99\text{m}}\text{Tc}$ -HMPAO = technetium-99m hexamethylpropylene amine oxime; ^{123}I -IMP = ^{123}I -N-isopropyl-iodoamphetamine; FDG = fluorodeoxyglucose; HCI = hypometabolic convergence; sc-SPM = single-case statistical parametric mapping; t-sum = PMOD Alzheimer discrimination analysis tool; TP = temporoparietal.

Table 1 Technical features of imaging metrics

	Acquisition	Automation	Stability	Intensity	Normative population	Abnormality threshold	Cost	Notes
Increased binding of amyloid imaging agents with PET								
Visual read	Any acquisition compliant with current guidelines	Fully subjective subject classification, no automation	IRR: 0.70–0.96	Software for image management required	Not required	Not required	Very low	
			TRTR: 0.89–0.99	20–30 s p/t per scan				
SUVr	Any acquisition compliant with current guidelines	Automated uptake quantification and subject classification	Not available	Software required for semiquantitative uptake estimates	Not required	1.3–1.5	Medium	
DVR	Any acquisition compliant with current guidelines	Automated uptake quantification and subject classification	Not available	Software required for semiquantitative uptake estimates	Not required	>1.2	Medium	
Temporoparietal hypometabolism on ¹⁸F-FDG-PET								
Visual read	Any acquisition compliant with ESNM/SNM guidelines	Fully subjective, no automation	IRR: 0.56–0.98	No apparel required	Not required	Not required	Very low	Poor stability
			TRTR: 0.78–0.97	20–30 s p/t per scan				
t-Sum	Any acquisition compliant with current guidelines	Fully automated	Not available	Commercial software required (PMOD software)	Healthy elders	t-sum >11,090	Low	
				Digital images usually in DICOM format				
				3 min p/t per scan				
HCI	Any acquisition compliant with current guidelines	Fully automated	Not available	Commercial software required (MATLAB for SPM)	Healthy elders	z Score of ≥ 1.5	Low	Age-correction still needs to be implemented
				Digital images required in NIFTI or DICOM format				
				3 min p/t per scan				
Neurostat/3D-SSP	Any acquisition compliant with current guidelines	Fully automated	Not available	Freely available software	Healthy elders	z Score of ≥ 1.5	Low	
				Digital images required in Analyze format				
				2 min p/t per scan				
sc-SPM	Any acquisition compliant with current guidelines	Fully automated	Not available	Digital images required in NIFTI or DICOM format	Healthy elders	z Score of ≥ 1.5	Low	
Temporoparietal hypoperfusion on SPECT								
Visual read	Any acquisition compliant with guidelines	Fully subjective, no automation	TRTR: 0.50–0.86	No apparel required	Not required	Not required	Very low	
			IRR: 0.76–0.89	20–30 s p/t per scan				
Semiquantitative/quantitative	Any acquisition compliant with current guidelines	Automated	Not available	Software for image management required	Healthy elders	z Score of ≥ 1.5	Low	

Continued

Table 1 Continued

	Acquisition	Automation	Stability	Intensivity	Normative population	Abnormality threshold	Cost	Notes
Medial temporal atrophy on MRI								
Visual read	T1-weighted acquisition	Fully subjective, no automation	TRTR: 0.82–0.97	No apparel required	Not required	0–1/2 + best at separating AD from healthy elders	Very low	Poorly sensitive to the mildest degrees of atrophy
			IRR: 0.82–0.86	20–30 s p/t per scan				
Manual segmentation	ADNI T1-weighted 3D	Fully manual	TRTR: 0.85–0.99	Software for image management required	ADNI healthy elders	Conventional, 95th percentile of age-specific normal distribution	Low once the segmentation apparel and procedure are set up	Normalization by head size not standardized
			IRR: 0.80–0.95	Digital images required in Analyze or DICOM format				
				1 h p/t per scan				
FreeSurfer	ADNI T1-weighted 3D	Fully automated	TRTR: 0.73–0.75	Freely available software required	ADNI healthy elders	Conventional, 95th percentile of age-specific normal distribution	Low	
				Digital images required in Analyze or NIFTI format				
				10 h p/t per scan				
AdaBoost-ACM	ADNI T1-weighted 3D	Fully automated	TRTR: 0.83–0.85	Freely available software required	ADNI healthy elders	Conventional, 95th percentile of age-specific normal distribution	Low	
				Digital images required in Analyze or NIFTI format				
				10 min p/t per scan				
BrainVISA/SACHA	ADNI T1-weighted 3D	Fully automated	TRTR: 0.91–0.98	Commercial software required (The Anatomist)	ADNI healthy elders	Conventional, 95th percentile of age-specific normal distribution	Low	
				Digital images required in NIFTI format				
Linear measure	T1-weighted acquisition	Fully manual	TRTR and IRR: 0.90–0.99	No apparel required	Not required	Conventional, 95th percentile of age-specific normal distribution	Very low	

Abbreviations: AD = Alzheimer disease; ADNI = Alzheimer's Disease Neuroimaging Initiative; DVR = distribution volume ratio; ESNM/SNM = European School of Nuclear Medicine/Society of Nuclear Medicine; FDG = fluorodeoxyglucose; HCI = hypometabolic convergence index; IRR = interrater reliability; sc-SPM = single-case statistical parametric mapping; SSP = stereotactic surface projection; SUVR = standardized uptake value ratio; 3D = 3-dimensional; TRTR = test-retest reliability; t-sum = PMOD Alzheimer discrimination analysis tool.

$(100 - \text{sensitivity})/\text{specificity}$. $LR+ \geq 5$ and $LR- \leq 0.2$ are generally regarded as clinically meaningful, i.e., diagnostically useful.³³ We analyzed separately the accuracy for the discrimination of persons with AD from healthy elderly subjects ("diagnostic," dementia stage) and for the discrimination of pMCI from npMCI ("prognostic," MCI stage).

To obtain pooled measures of sensitivity and specificity, we used a classic Bayesian approach.³⁴ (Details are provided in the supplementary material on the *Neurology*[®] Web site at www.neurology.org.) The estimation was repeated for each set of studies that investigated the same operating metrics on the same type of diagnostic groups.

Statistical analyses were performed with SPSS 12.0.1 (SPSS Inc., Chicago, IL) using 1-way analysis of variance (ANOVA)

and nested ANOVA to test whether diagnostic and prognostic $LR+$ and $LR-$ variability was attributable to differences among markers, metrics, and submarkers or attributable to variability among the metrics within markers, among the metrics within submarkers, or among the submarkers within markers. Statistical analyses and plots were restricted to metrics used by at least 3 studies. Through a linear regression analysis, we investigated the effect of age, disease severity, group size, and follow-up duration on sensitivity, specificity, and LR values.

RESULTS Table 2 shows sensitivity and specificity values pooled across markers, submarkers, and metrics. Diagnostic accuracy was highest for amyloid imaging

Table 2 Accuracy figures of imaging markers for AD at the dementia and MCI stages^a

	Specificity				Sensitivity				e-Ref.
	Dementia stage		MCI stage		Dementia stage		MCI stage		
	% (95% CI)	No. of healthy elders (no. of studies)	% (95% CI)	No. with npMCI (no. of studies)	% (95% CI)	No. with AD (no. of studies)	% (95% CI)	No. with pMCI (no. of studies)	
Increased binding of amyloid imaging agents with PET									
All metrics	85 (81-88)	510 (19)	56 (49-64)	134 (8)	88 (84-91)	453 (19)	82 (74-88)	101 (8)	
¹¹ C-PiB									
Visual read	85 (64-95)	20 ^b (1)	NA		100 (84-100)	21 ^b (1)	NA		e1
SUVR	83 (76-88)	136 (6)	53 (44-61)	106 (6)	91 (84-95)	104 (6)	82 (74-88)	87 (6)	e2-e12
DVR	88 (77-95)	32 (3)	56 (34-75)	18 ^b (1)	93 (85-98)	32 (3)	100 (52-100)	5 ^b (1)	e13, e14
¹⁸ F ligands									
SUVR	86 (81-90)	322 (9)	80 (48-94)	10 ^b (1)	87 (83-91)	296 (9)	78 (46-93)	9 ^b (1)	e15-e23
Temporoparietal hypometabolism on ¹⁸ F-FDG-PET									
All metrics	84 (81-87)	1011 (21)	74 (68-78)	291 (10)	86 (84-89)	1897 (37)	76 (70-82)	241 (10)	
Neurostat/3D-SSP	90 (85-94)	303 (9)	76 (60-89)	28 ^b (2)	88 (84-91)	686 (12)	77 (60-90)	22 ^b (2)	e24-e36
t-sum/HCI	85 (79-90)	266 (4)	55 (47-64)	158 (3)	87 (84-90)	743 (4)	69 (57-78)	65 (3)	e37-e42
sc-SPM	83 (77-88)	376 (5)	92 (81-97)	53 (3)	84 (78-90)	331 (6)	72 (61-84)	49 (3)	e4, e13, e43-e49
Visual read	68 (57-78)	66 (3)	74 (60-84)	52 ^b (2)	85 (78-90)	137 (5)	94 (88-97)	105 ^b (2)	e1, e13, e35, e50-e53
Temporoparietal hypoperfusion on SPECT									
All metrics	84 (81-87)	773 (28)	64 (55-72)	137 (5)	76 (74-79)	1,268 (32)	78 (72-85)	166 (6)	
^{99m} Tc-ECD and ¹²³ I-IMP									
Visual read	100 (71-100)	9 ^b (1)	NA		70 (45-88)	14 ^b (1)	NA		e28
Semiquantitative/ quantitative	92 (66-98)	13 ^b (1)	58 (46-70)	54 (3)	89 (84-92)	231 ^b (2)	79 (71-91)	121 (4)	e30, e54-e57
^{99m} Tc-HMPAO									
Visual read	84 (78-88)	344 (11)	NA		68 (63-72)	422 (11)	NA		e58-e69
Semiquantitative/ quantitative	83 (79-87)	407 (15)	64 (51-75)	83 ^b (2)	81 (78-84)	601 (18)	78 (65-88)	45 ^b (2)	e57, e60, e61, e70-e85
Medial temporal atrophy on MRI									
All metrics	81 (79-82)	2,687 (66)	73 (69-76)	1,196 (26)	75 (73-77)	2,304 (72)	62 (58-66)	592 (27)	
Temporal horn									
Linear measure	87 (82-91)	142 (7)	100 (63-100)	8 ^b (1)	70 (63-76)	151 (7)	40 (17-69)	10 ^b (1)	e86-e90
Amygdala									
Visual read	88 (75-95)	40 ^b (1)	NA		69 (58-78)	77 ^b (1)	NA		e91
Manual segmentation	77 (69-83)	79 (8)	NA		79 (73-84)	155 (10)	NA		e92-e98
Hippocampus									
Visual read	79 (75-83)	616 (10)	75 (67-82)	245 (6)	70 (65-74)	526 (12)	60 (51-68)	171 (7)	e1, e30, e53, e87, e90, e91, e99-e108
Manual segmentation	82 (78-85)	565 (18)	81 (73-87)	264 (5)	79 (76-82)	528 (20)	58 (47-68)	85 (5)	e86, e88, e90, e92, e93, e95, e96, e100, e109-e121
FreeSurfer/ AdaBoost-ACM/ BrainVISA	81 (77-85)	519 (7)	66 (61-71)	247 (5)	72 (67-77)	430 (7)	70 (63-76)	136 (5)	e47, e122-e128

Continued

Table 2 Continued

	Specificity				Sensitivity				e-Ref.
	Dementia stage		MCI stage		Dementia stage		MCI stage		
	% (95% CI)	No. of healthy elders (no. of studies)	% (95% CI)	No. with npMCI (no. of studies)	% (95% CI)	No. with AD (no. of studies)	% (95% CI)	No. with pMCI (no. of studies)	
Entorhinal cortex									
Visual read	86 (81–90)	379 (3)	78 (54–91)	18 ^b (1)	58 (49–68)	132 (3)	86 (56–97)	11 ^b (1)	e87, e91, e106, e108
Manual segmentation	76 (70–81)	216 (9)	72 (61–83)	234 (4)	83 (78–88)	194 (9)	66 (56–76)	80 (4)	e96, e101, e109, e110, e115, e116, e119–e121, e129, e130
FreeSurfer	85 (79–89)	101 ^b (2)	66 (59–72)	180 (4)	71 (64–77)	81 ^b (2)	57 (49–66)	99 (4)	e122, e123, e126
Linear measure	80 (63–90)	30 ^b (1)	NA		67 (49–81)	30 ^b (1)	NA		e120

Abbreviations: AD = Alzheimer disease; CI = confidence interval; DVR = distribution volume ratio; FDG = fluorodeoxyglucose; HCI = hypometabolic convergence index; MCI = mild cognitive impairment; NA = not available; npMCI = nonprogressed MCI; pMCI = progressed MCI; sc-SPM = single-case statistical parametric mapping; SUVR = standardized uptake value ratio; ^{99m}Tc-HMPAO = technetium-99m hexamethylpropylene amine oxide; t-sum = PMOD Alzheimer discrimination analysis tool; 3D-SSP = 3-dimensional stereotactic surface projection; ¹¹C-PiB = ¹¹C-labeled Pittsburgh compound B; ^{99m}Tc-ECD = technetium-99m bicitate; ¹²³I-IMP = ¹²³I-N-isopropyl-iodoamphetamine.

^aDementia stage denotes accuracy for the discrimination of patients with AD dementia from healthy elders (diagnostic accuracy). MCI stage denotes accuracy for the discrimination of patients with MCI who subsequently progressed to AD from patients with MCI who did not progress (prognostic accuracy). Figures denote pooled mean values and 95% CIs weighted for group size and computed through a Bayesian approach.³⁴ NA: no published studies on this specific metric. The e-references are listed in the supplementary material.

^bDenotes metrics with <3 studies.

and progressively lower for ¹⁸F-FDG-PET, SPECT, and MRI. Prognostic accuracy had a similar pattern across markers, but was generally lower than diagnostic accuracy.

LR analysis, dementia stage. The analysis of LR+ (figure 2A) mirrored the accuracy analysis pattern; it was best for amyloid imaging (9.4) and poorest for MRI (4.4). Considering amyloid imaging submarkers, LR+ values were best for ¹⁸F ligands, whereas for MRI submarkers, the best were for temporal horn and the poorest for entorhinal cortex.

At the metrics level, the variability of LR+ was often as high as between markers, in particular for ¹⁸F-FDG-PET (range: 13.3–2.4), and for MRI (10.7–4.2). The variability among metrics was lower for the other 2 markers.

In LR– (figure 2B), for markers, the best values were for amyloid imaging (0.08) and the poorest for MRI (0.25). LR– values across amyloid imaging submarkers were rather homogeneous and little variation was also detected for MRI submarkers.

The variability of LR– of metrics was much higher, especially for amyloid imaging (0.01–0.10), and for ¹⁸F-FDG-PET (0.05–0.23). Variability among metrics was lower for the other 2 markers: 0.21 to 0.32 (MRI metrics), and 0.13 to 0.17 (SPECT metrics). For detailed information, see table e-1.

The variability of LR+ within metrics was even greater than across markers and metrics. Many metrics spanned 2 orders of magnitude, LR+ ranging

from the poorest values between 1 and 3 up to between 70 and 100 (figure 2A). The variability of LR– within metrics was similar, spanning 2 orders of magnitude from 0.01 to 1.00 (figure 2B).

LR analysis, MCI stage. Prognostic was generally poorer than diagnostic LR+ figures, being more than 5 for only ¹⁸F-FDG-PET (7.5). The pattern was also different; the second best LR+ value was that of MRI (2.6), followed by SPECT (2.2), and by amyloid imaging (1.7). It should be noted, however, that the number of studies contributing to prognostic LR+ estimation was much lower than that of diagnostic LR+. Considering MRI submarkers, prognostic LR+ was 2.9 for hippocampus and 2.2 for entorhinal cortex (figure 3A).

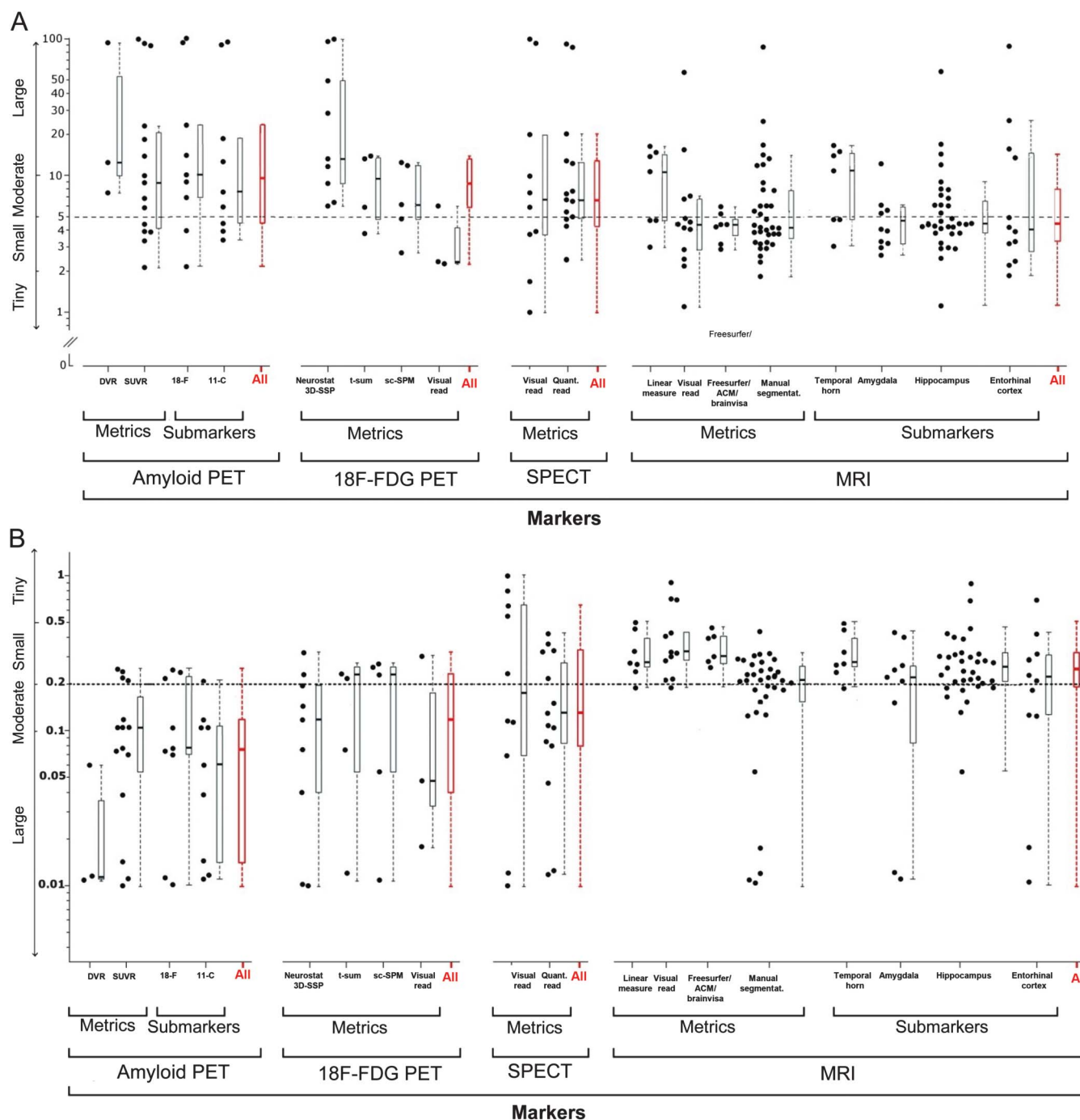
In analogy with the pattern of diagnostic LR+, the variability across metrics was in some cases at least as large as that across markers. Prognostic LR+ of ¹⁸F-FDG-PET metrics ranged from 12.8 to 1.7. The variability across MRI metrics was lower (3.2–1.8).

For markers, LR– values (figure 3B) were best for amyloid imaging (0.11) and poorest for MRI (0.49) and ¹⁸F-FDG-PET (0.50). For MRI submarkers, LR– was 0.49 for hippocampus and 0.56 for entorhinal cortex.

Again, LR– values of ¹⁸F-FDG-PET metrics were quite heterogeneous (0.08–0.64), whereas the variability across MRI metrics was lower (0.46–0.50). For details, see table e-1.

The variability of prognostic LR+ within metrics spanned 1 order of magnitude, with few exceptions spanning 2 orders of magnitude (from approximately

Figure 2 Diagnostic (A) positive and (B) negative likelihood ratio (LR+ and LR−) for correct classification between patients with Alzheimer disease and healthy subjects broken down by markers by metrics and by markers by submarkers



Only metrics with at least 3 studies are shown. Boxplots denote median, first, and third quartiles, and whiskers denote minimum and maximum values (excluding outliers). Dashed lines indicate the conventional thresholds of clinical relevance (5 for LR+ and 0.2 for LR−). DVR = distribution volume ratio; FDG = fluorodeoxyglucose; sc-SPM = single-case statistical parametric mapping; SUVR = standardized uptake value ratio; 3D-SSP = 3-dimensional stereotactic surface projection.

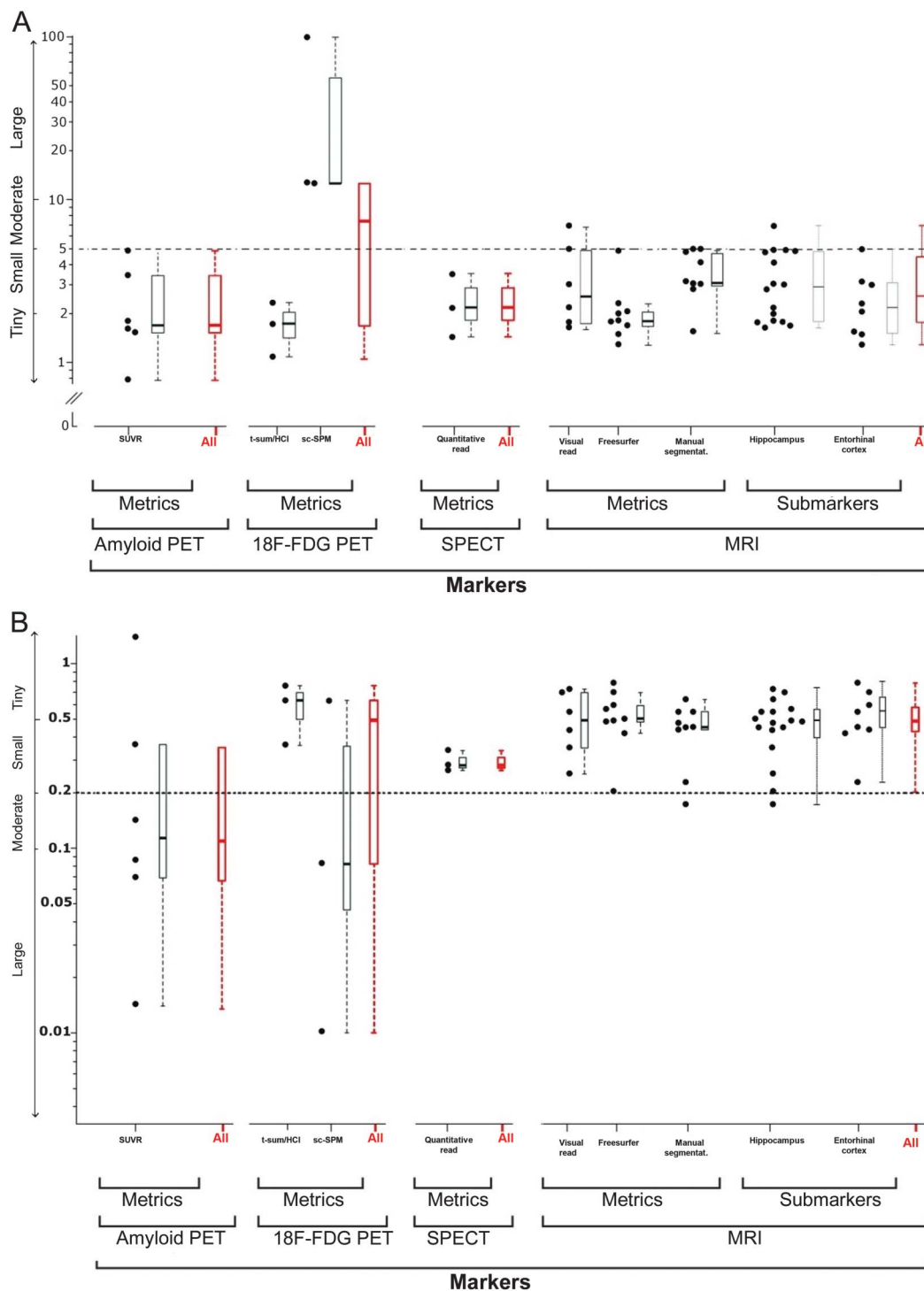
1 to 10). The variability of LR− within metrics was similar, with a few exceptions spanning 2 orders of magnitude (¹⁸F-FDG-PET sc-SPM and amyloid PET-standardized uptake value ratio) (figure 3B).

Proportion of explained variance of LR estimates, dementia stage. Markers accounted for 11% of LR+ and 24% of

LR− variance and metrics for 13% and 29%, respectively (figure 4A). When markers were divided into “functional” (¹⁸F-FDG-PET and SPECT) and “structural” (MRI), they accounted for 12% of LR+ variance. Of all metrics, those with the largest variability were ¹⁸F-FDG-PET metrics (39%) for LR+, and MRI metrics (37%) for LR−. The variance of LR− explained by

Figure 3

Prognostic (A) positive and (B) negative likelihood ratio (LR+ and LR–) for correct classification of patients with progressed vs nonprogressed mild cognitive impairment, broken down by markers by metrics and by markers by submarkers

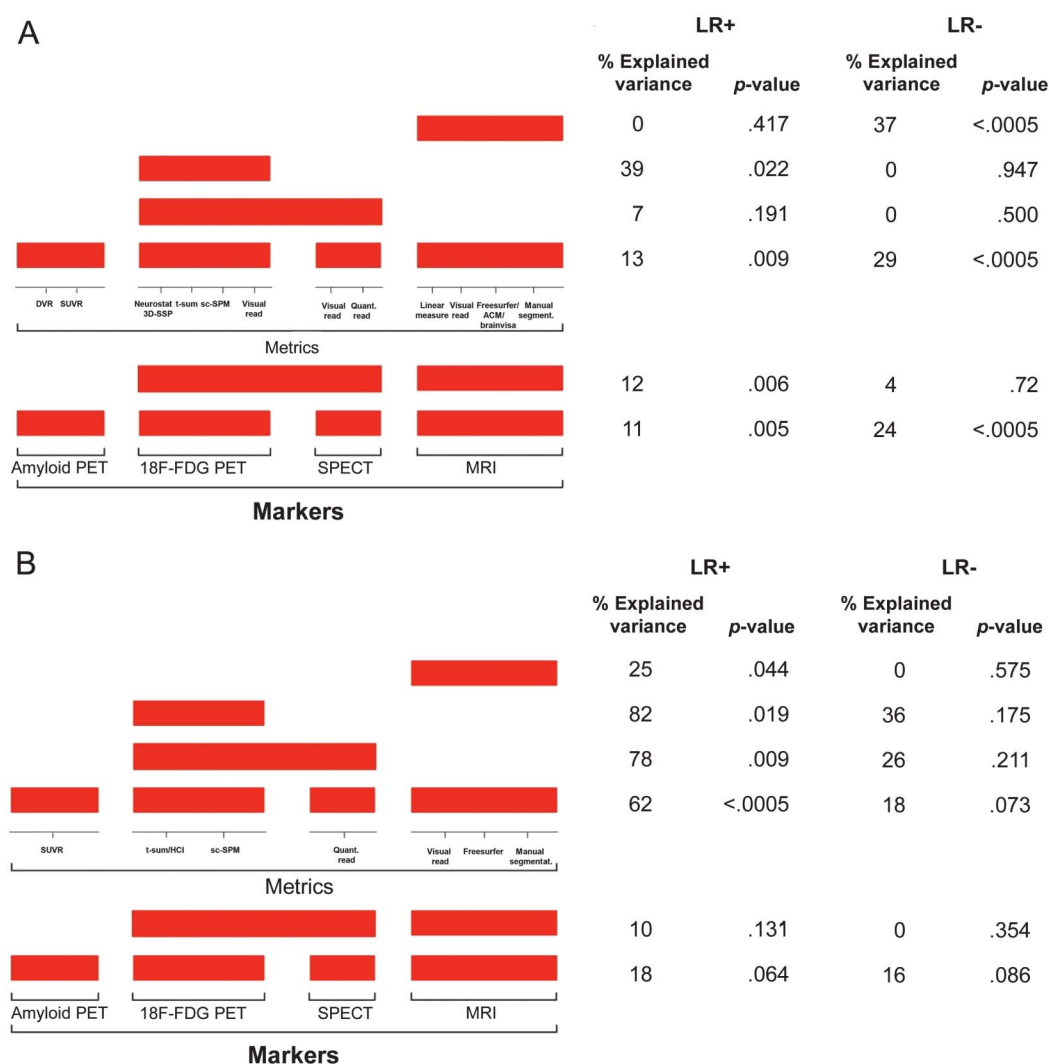


Only metrics with at least 3 studies are shown. Boxplots denote median, first, and third quartiles, and whiskers denote minimum and maximum values (excluding outliers). Dashed lines indicate the conventional thresholds of clinical relevance (5 for LR+ and 0.2 for LR–). FDG = fluorodeoxyglucose; HCI = hypometabolic convergence index; sc-SPM = single-case statistical parametric mapping; SUVR = standardized uptake value ratio.

metrics remained significant even when tested with the more stringent nested ANOVA within markers (17%) and submarkers (15%). When restricted to MRI, nested ANOVA analysis showed that metrics within MRI submarkers accounted for 41% of diagnostic LR– variance.

Proportion of explained variance of LR estimates, MCI stage. When compared with diagnostic LR+ variance, both markers and metrics accounted for a larger proportion of prognostic variance (18% and 62%, respectively) (figure 4B). In contrast, compared with

Figure 4 Proportion of explained variance and significance of positive and negative likelihood ratio (LR+ and LR−) for correct classification between (A) patients with Alzheimer disease and healthy subjects, and (B) patients with progressed and nonprogressed mild cognitive impairment



Analysis included only values for metrics with at least 3 studies. Rectangles denote the markers or metrics whose LR variance is computed. DVR = distribution volume ratio; FDG = fluorodeoxyglucose; sc-SPM = single-case statistical parametric mapping; SUVR = standardized uptake value ratio; 3D-SSP = 3-dimensional stereotactic surface projection.

diagnostic LR− variance, markers and metrics accounted for a lower proportion of prognostic variance, 16% and 18%, respectively.

Similarly to diagnostic metrics, the prognostic metrics with the largest LR+ variability were ¹⁸F-FDG-PET metrics (82%). Metrics accounted for 25% of prognostic LR+ variance of the MRI marker. When considered together, SPECT and ¹⁸F-FDG-PET metrics accounted for 78% of prognostic LR+ variance.

The prognostic variance of metrics remained significant even when tested with nested ANOVA within markers (68%). When restricted to SPECT and ¹⁸F-FDG-PET metrics, nested ANOVA analysis showed that these functional metrics accounted for 86% of prognostic variance.

Effect of confounders on LR estimates. Specific analyses regarding the effect of study group size, follow-up duration, age, and disease severity on accuracy figures are reported in the supplementary material.

DISCUSSION We have estimated diagnostic and prognostic accuracy of different AD markers as well as pertinent metrics, and the amount and source of variance among them. We have shown that the diagnostic and prognostic accuracy of imaging AD biomarkers is at least as dependent on how the biomarker is measured as on the type of biomarker itself. While acknowledging that imaging biomarkers capture different neurobiological constructs (brain amyloidosis, neuronal injury at the molecular level, and neuronal injury at the gross

structural level), this observation provides empirical support to current efforts aimed at developing standard operating procedures (SOPs) for AD biomarkers.^{7,35} Such efforts are key to the use of imaging biomarkers in the diagnostic routine and in clinical trials.

Diagnostic LR_s were generally better than prognostic LR_s: diagnostic LR_s+ were approximately >5 for all markers and metrics (except ¹⁸F-FDG-PET visual read), and diagnostic LR_s− were generally <0.20, except for MRI metrics. This is expected in that biological changes in patients with pMCI are milder than in patients with AD dementia.^{36–41} The increasing awareness of AD and options for early diagnosis make biomarkers particularly useful in clinical practice to distinguish pMCI from npMCI. Here, LR_s+ were <3 for all metrics and markers, except ¹⁸F-FDG-PET; the pattern was similar for LR_s−, where all markers and the majority of metrics yielded LR_s− >0.45, except amyloid imaging. Alternatively, better LR_s of diagnostic studies might be attributable to cross-sectional case-control studies yielding optimistic estimates of sensitivity and specificity.⁴²

LR point estimates of amyloid imaging metrics tended to be better than the other metrics. This is in line with the current understanding of the AD pathophysiology, positing that brain amyloidosis is a necessary condition for AD-related neurodegeneration to take place.^{4,43} On the contrary, LR_s+ and LR_s− figures were the poorest for MRI metrics in almost all conditions. This is expected in view of the little specificity of medial temporal atrophy, which is featured in AD as well as in a proportion of cognitively healthy older persons.⁴⁴ It should be emphasized, however, that because of limitations of the current review, we cannot conclude that a metric or a marker is better than another for clinical use. For instance, the number of studies with amyloid imaging is by far lower than those with MRI and, in the prognostic condition, also than those with ¹⁸F-FDG-PET. More amyloid imaging studies, possibly focused on differential diagnosis, are needed to consolidate the pertinent estimates on LR_s and to allow comparisons among different tracers.

The metrics with the largest variability were those of ¹⁸F-FDG-PET. Interestingly, diagnostic LR_s+ for t-sum was better than sc-SPM, and vice versa for prognostic LR_s+. This attests to the benefits of standardizing ¹⁸F-FDG-PET metrics.

Importantly, if the variability of diagnostic LR_s across metrics varied by 1 order of magnitude (i.e., 10-fold), the variability within a metric varied by as many as 2 orders of magnitude (i.e., 100-fold). The within-metric variability of prognostic LR_s also varied by 1 order of magnitude. This observation militates in favor of standardization of metrics, which should reduce this 100-fold variability to close to zero.

Metrics vary for a number of features such as dependency on i) a specific (sometimes nonroutine) image acquisition protocol and ii) a human rater and automation; iii) stability over time (test-retest reliability) and across raters (interrater reliability); iv) feasibility in routine clinical settings, where human and technological resources are tailored to the use of routine tests; availability of v) rigorously standardized operating procedures for measurement vi) of a reference normative population and vii) of reliable abnormality thresholds; and viii) cost of the overall acquisition and measurement procedure. All of the above issues should be addressed by standardization efforts for metrics to be adopted in the clinical routine and to be used as the reference for validation of automated algorithms. The practical message to clinical neurologists is that using AD markers in the diagnostic pathway of patients with cognitive impairment is not a guarantee of greater accuracy per se. Because accuracy largely depends on how a marker is analyzed, clinicians wishing to follow the International Working Group or NIA-AA diagnostic criteria can i) use metrics for which SOPs are available and whose accuracy is known (e.g., FreeSurfer/NeuroQuant for medial temporal atrophy or ¹⁸F-florbetapir for cortical amyloid burden), ii) empirically measure in their own setting the accuracy of the metric they wish to use, or iii) wait for SOPs to be developed for other metrics.

This review has a number of limitations that should be noted. Because of the small number of studies, we did not address the accuracy of imaging biomarkers for the differential diagnosis of dementia type (AD vs Lewy body dementia, vs frontotemporal degeneration, etc.). With the hopeful advent of drugs affecting specific core pathophysiologic substrates of AD, this issue may become of greater relevance and need to be properly addressed, also because differential diagnosis is crucial in the clinical practice. However, we found that how a marker is measured is as relevant as which marker is considered even for a less “clinically relevant” and “easier” comparison (AD vs healthy), further reinforcing the need of standardized measurement of biomarkers.

We accepted the definitions of AD and MCI adopted by the reviewed studies, including exclusion criteria (e.g., vascular disease, medications), thus accepting the inherent clinical heterogeneity, which may be enhanced by the fact that some patients were from research cohorts. We did not consider neuropathologic diagnosis of AD because few studies have histopathologic confirmation of AD diagnosis and we recognized that this is a limitation of our review.

Data regarding the “classification” of AD dementia vs cognitively normal elders is only a necessary but not sufficient indicator of a test’s value and does not reflect its diagnostic accuracy in clinical settings. Further tests of “diagnostic” value would be those that help in the differential diagnosis (e.g., among different dementia

cases) for predicting postmortem neuropathology and a person's clinical course (as in the MCI analysis), and eventually, for predicting response to treatment. Another limitation pertains to the use of imaging markers for prognosis⁴ and differential diagnosis⁵ in the context of the new criteria. Here, imaging needs to be used together with biological (CSF) markers, and the contribution of biological markers to LR of imaging markers will need to be investigated in future studies with pathologic confirmation.

The definition we chose to classify metrics should be taken with due caution: neurodegenerative changes in the medial temporal lobe have largely been assessed using volumetric MRI, but evidence showed that they can also be accurately studied with FDG-PET.⁴⁵ Confounders had little effect on diagnostic and prognostic accuracy values, with the exception of the positive association of age with LR− and its negative association with specificity in AD, and negative association of age with LR+ and specificity in MCI stage. We believe that this observation is attributable to the relatively higher frequency of abnormal markers in elderly persons despite no disease.⁴⁶ For the MCI stage, there was a significant effect of study group size on LR−, indicating a slight increase of false-negative rates for a given true-positive rate with increasing size. This is understandable in light of observations that smaller studies frequently show better accuracy values because of stricter selection of cases. Lastly, it should be noted that some sources of variability (e.g., ethnicity, application of diagnostic criteria, inclusion and exclusion criteria, case mix, socioeconomic status) might have escaped our analyses because of the intrinsic limitations of this type of meta-analysis.

AUTHOR CONTRIBUTIONS

Drafting/revising the manuscript for content: Frisoni, Bocchetta, Pasqualetti, Chételat, Rabinovici, Herholz, Kaye, Jack, Rowe, Jagust, Wahlund, Brooks, Nordberg, Scheltens, Reiman, Weiner, de Leon. Study concept or design: Frisoni, DeCarli, Barkhof, Herholz, Kaye, Jack, Rowe, Jagust, Wahlund, Brooks, Nordberg, Scheltens, Reiman, Fox, Black, Sperling, Johnson, Weiner, Carrillo, Thies. Analysis or interpretation of data: Frisoni, Bocchetta, Pasqualetti. Statistical analysis: Bocchetta, Pasqualetti. Study supervision or coordination: Frisoni.

ACKNOWLEDGMENT

The authors thank Marco Lorenzi, PhD, for his valuable help with data analysis.

STUDY FUNDING

This study was partially funded by the Alzheimer's Association grant IIRG-10-174022, "A Harmonized Protocol for Hippocampal Volumetry: An EADC-ADNI Effort."

DISCLOSURE

G. Frisoni has served on advisory boards for Lilly, BMS, Bayer, Lundbeck, Elan, AstraZeneca, Pfizer, TauRx, Wyeth, and GE; he is a member of the editorial boards of *Lancet Neurology*, *Aging Clinical and Experimental Research*, *Alzheimer Disease & Associated Disorders*, *Neurodegenerative Diseases*, and Imaging Section Editor of *Neurobiology of Aging*; he has

received grants from Wyeth International, Lilly International, Lundbeck Italia, GE International, Avid/Lilly, and the Alzheimer's Association. M. Bocchetta, G. Chételat, and G. Rabinovici report no disclosures. M. de Leon has served on the scientific advisory board for Roche; he is a holder of image analysis patents through New York University. J. Kaye received research support from the Department of Veterans Affairs and the NIH; individuals that work in the research centers he directs received research support from Johnson & Johnson, Roche, and Bristol-Myers Squibb. J. Kaye was compensated for serving on a data monitoring committee for Eli Lilly, and as a paid advisor for Janssen Pharmaceutical; he received reimbursement through Medicare or commercial insurance plans for providing clinical assessment and care for patients; he has been salaried to see patients at the Portland VA Medical Center; he served as an unpaid Vice-Chair for the International Professional Interest Area Work Group of the ISTAART and as an unpaid Commissioner for the Center for Aging Services and Technologies; he serves on the editorial advisory board of the journals *Alzheimer's & Dementia* and *Frontiers of Aging Neuroscience*. E. Reiman served as a scientific advisor to Sygnis, AstraZeneca, Bayer, Eisai, Elan, Eli Lilly, GlaxoSmithKline, Intellect, Link Medicine, Novartis, Siemens, and Takeda; he has had research contracts with AstraZeneca and Avid/Eli Lilly; a patent pending for a biomarker strategy to evaluate preclinical AD treatments (through Banner Health); and research grants from the National Institute on Aging, Anonymous Foundation, Nomis Foundation, Banner Alzheimer's Foundation, and the State of Arizona. P. Scheltens serves on the advisory boards of Genentech, Novartis, Pfizer, Roche, Danone, Nutricia, Jansen AI, Baxter, and Lundbeck; he has been a speaker at symposia organized by Lundbeck, Lilly, Merz, Pfizer, Jansen AI, Danone, Novartis, Roche, and Genentech; he serves on the editorial board of *Alzheimer's Research & Therapy* and *Alzheimer Disease & Associated Disorders*; he is a member of the scientific advisory board of the European Union Joint Programming Initiative and the French National Plan Alzheimer. The Alzheimer Center receives unrestricted funding from various sources through the VUmc Fonds; he receives no personal compensation for the activities mentioned above. F. Barkhof reports no disclosures. S. Black has received funding in the past 2 years for ad hoc consulting from Pfizer, Novartis, Roche, Bristol-Myers Squibb, GlaxoSmithKline, and Elan. She has received speaker's honoraria for CME from Pfizer, Novartis, and Eisai. Dr. Black's unit has received contract research funds from GlaxoSmithKline, Roche, Pfizer, and Elan and research funds from the Canadian Institutes of Health Research (MOP-13129, MOP-106485, MOP-82744), NIH (ADNI), Heart and Stroke Foundation Centre for Stroke Recovery, Heart and Stroke Foundation of Canada (T6075, T6383), Alzheimer's Drug Discovery Foundation, W. Garfield Weston Foundation, and Brain Canada. She has received salary support from the Brill Chair in Neurology, the Sunnybrook Research Institute, and the Department of Medicine, University of Toronto. D. Brooks and M. Carrillo report no disclosures. N. Fox holds a patent for QA Box that may accrue revenue. In the last 2 years, his research group has received payment for consultancy or for conducting studies from AVID, Bristol-Myers Squibb, Elan Pharmaceuticals, Eisai, Lilly Research Laboratories, GE Healthcare, IXICO, Janssen Alzheimer Immunotherapy, Johnson & Johnson, Janssen-Cilag, Lundbeck, Neurochem Inc., Pfizer Inc., Sanofi-Aventis, and Wyeth Pharmaceuticals. He receives research support from MRC (G0801306 [PI], G0601846 [PI]), NIH (U01 AG024904 (coinvestigator; subcontract), Alzheimer Research Trust (ART/RF/2007/1 [PI]), and NIHR (senior investigator). K. Herholz reports no disclosures. A. Nordberg has been the PI for clinical trials sponsored by TorreyPines Therapeutics, GSK, Wyeth, and Bayer Pharma; she served on the advisory board for Elan, Pfizer, GSK, Novartis, Lundbeck AB, Johnson & Johnson, GE Healthcare, and Avid; she received honorarium for lectures from Novartis, Janssen-Cilag, Pfizer, and Merck, and research grants from Novartis, Pfizer, GE Healthcare, Johnson & Johnson, and Bayer Pharma; she owns no stocks and is a member of the editorial advisory board for *Current Alzheimer Research*, *Journal of Alzheimer's Disease*, and *Alzheimer's Research & Therapy*. C. Jack serves as a consultant for Janssen, Bristol-Myers Squibb, General Electric, Siemens, and Johnson & Johnson, and is involved in clinical trials sponsored by Allon and Baxter, Inc.; he receives research funding from the NIH, and the Alexander Family Alzheimer's Disease Research Professorship of the Mayo Foundation. W. Jagust has served as a consultant to Siemens, Genentech, TauRx, and Janssen Alzheimer Immunotherapy; he receives research support from NIH (AG034570, AG025303). K. Johnson and C. Rowe report no disclosures.

R. Sperling has served as a paid consultant for Bayer, Biogen Idec, Bristol-Myers Squibb, Eisai, Janssen Alzheimer Immunotherapy, Pfizer, Merck, Roche, Satori, and as an unpaid consultant to Avid; she is a site coinvestigator for Avid, Bristol-Myers Squibb, Pfizer, and Janssen Alzheimer Immunotherapy clinical trials. She has spoken at symposia sponsored by Eli Lilly, Pfizer, and Janssen Alzheimer Immunotherapy. W. Thies and L. Wahlund report no disclosures. M. Weiner served on the scientific advisory board for Lilly, Araclon and Institut Catala de Neurociencies Aplicades, Gulf War Veterans Illnesses Advisory Committee, VACO, Biogen Idec, Pfizer, and BOLT International; he is a consultant for AstraZeneca, Araclon, Medivation/Pfizer, Ipsen, TauRx Therapeutics Ltd., Bayer Healthcare, Biogen Idec, ExonHit Therapeutics, SA, Servier, Synarc, Pfizer, Janssen, Harvard University, and KLJ Associates; he received funds for travel from NeuroVigil, Inc., CHRU-Hopital Roger Salengro, Siemens, AstraZeneca, Geneva University Hospitals, Lilly, University of California, San Diego-ADNI, Paris University, Institut Catala de Neurociencies Aplicades, University of New Mexico School of Medicine, Ipsen, CTAD (Clinical Trials on Alzheimer's Disease), Pfizer, AD PD meeting, Paul Sabatier University, Novartis, Tohoku University, Fundacio ACE, and Travel eDreams, Inc.; he is a member of the editorial advisory board of *Alzheimer's & Dementia* and *MRI*; he received honoraria from NeuroVigil, Inc., Institut Catala de Neurociencies Aplicades, PMDA/Japanese Ministry of Health, Labour, and Welfare, Tohoku University, and Alzheimer's Drug Discovery Foundation; he received research support from commercial (Merck and Avid) and government (DOD and VA) entities; he holds stock options from Synarc and Elan; he received funds from organizations contributing to the Foundation for NIH and thus to the NIA-funded Alzheimer's Disease Neuroimaging Initiative: Abbott, Alzheimer's Association, Alzheimer's Drug Discovery Foundation, Anonymous Foundation, AstraZeneca, Bayer Healthcare, BioClinica, Inc. (ADNI 2), Bristol-Myers Squibb, Cure Alzheimer's Fund, Eisai, Elan, Gene Network Sciences, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson & Johnson, Eli Lilly & Company, Medpace, Merck, Novartis, Pfizer Inc., Roche, Schering-Plough, Synarc, and Wyeth. P. Pasqualetti and C. DeCarli report no disclosures. Go to Neurology.org for full disclosures.

Received November 20, 2012. Accepted in final form April 12, 2013.

REFERENCES

- Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol* 2007;6:734–746.
- Dubois B, Feldman HH, Jacova C, et al. Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol* 2010;9:1118–1127.
- Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease. *Alzheimers Dement* 2011;7:280–292.
- Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease. *Alzheimers Dement* 2011;7:270–279.
- McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease. *Alzheimers Dement* 2011;7:263–269.
- Khachaturian ZS. Revised criteria for diagnosis of Alzheimer's disease: National Institute on Aging–Alzheimer's Association Diagnostic Guidelines for Alzheimer's Disease. *Alzheimers Dement* 2011;7:253–256.
- Frisoni GB, Hampel H, O'Brien JT, Ritchie K, Winblad B. Revised criteria for Alzheimer's disease: what are the lessons for clinicians? *Lancet Neurol* 2011;10:598–601.
- Frisoni GB, Winblad B, O'Brien JT. Revised NIA-AA criteria for the diagnosis of Alzheimer's disease: a step forward but not yet ready for widespread clinical use. *Int Psychogeriatr* 2011;23:1191–1196.
- Gauthier S, Patterson C, Gordon M, Soucy JP, Schubert F, Leuzy A. Commentary on "Recommendations from the National Institute on Aging–Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease." A Canadian perspective. *Alzheimers Dement* 2011;7:330–332.
- Clark CM, Pontecorvo MJ, Beach TG, et al. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: a prospective cohort study. *Lancet Neurol* 2012;11:669–678.
- Yang L, Rieves D, Ganley C. Brain amyloid imaging: FDA approval of florbetapir F18 injection. *N Engl J Med* 2012;367:885–887.
- European Medicines Agency. Amyvid (florbetapir 18F). EMA/696925/2012, EMA/H/C/0022422; January 2013.
- Committee for Medicinal Products for Human Use (CHMP). Qualification opinion of low hippocampal volume (atrophy) by MRI for use in clinical trials for regulatory purpose: in pre-dementia stage of Alzheimer's disease. EMA/CHMP/SAWP/809208/2011; November 17, 2011.
- Committee for Medicinal Products for Human Use (CHMP). Qualification opinion of Alzheimer's disease novel methodologies/biomarkers for PET amyloid imaging (positive/negative) as a biomarker for enrichment, for use in regulatory clinical trials in predementia Alzheimer's disease. EMA/CHMP/SAWP/892998/2011; February 16, 2012.
- Bloudek LM, Spackman DE, Blankenburg M, Sullivan SD. Review and meta-analysis of biomarkers and diagnostic imaging in Alzheimer's disease. *J Alzheimers Dis* 2011;26:627–645.
- Bohnen NI, Djang DS, Herholz K, Anzai Y, Minoshima S. Effectiveness and safety of 18F-FDG PET in the evaluation of dementia: a review of the recent literature. *J Nucl Med* 2012;53:59–71.
- Chetelat G, Baron JC. Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. *Neuroimage* 2003;18:525–541.
- Devous MD Sr. Functional brain imaging in the dementias: role in early detection, differential diagnosis, and longitudinal studies. *Eur J Nucl Med Mol Imaging* 2002;29:1685–1696.
- Dougall NJ, Bruggink S, Ebmeier KP. Systematic review of the diagnostic accuracy of ^{99m}Tc-HMPAO-SPECT in dementia. *Am J Geriatr Psychiatry* 2004;12:554–570.
- Herholz K, Ebmeier K. Clinical amyloid imaging in Alzheimer's disease. *Lancet Neurol* 2011;10:667–670.
- Jack CR Jr. Alliance for Aging Research AD Biomarkers Work Group: structural MRI. *Neurobiol Aging* 2011;32(suppl 1):S48–S57.
- Laforce R Jr, Rabinovici GD. Amyloid imaging in the differential diagnosis of dementia: review and potential clinical applications. *Alzheimers Res Ther* 2011;3:31.
- Mosconi L. Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease: FDG-PET studies in MCI and AD. *Eur J Nucl Med Mol Imaging* 2005;32:486–510.
- Patwardhan MB, McCrory DC, Matchar DB, Samsa GP, Rutschmann OT. Alzheimer disease: operating characteristics of PET—a meta-analysis. *Radiology* 2004;231:73–80.

25. Yuan Y, Gu ZX, Wei WS. Fluorodeoxyglucose-positron-emission tomography, single-photon emission tomography, and structural MR imaging for prediction of rapid conversion to Alzheimer disease in patients with mild cognitive impairment: a meta-analysis. *AJNR Am J Neuroradiol* 2009;30:404–410.
26. Zhang S, Han D, Tan X, Feng J, Guo Y, Ding Y. Diagnostic accuracy of (18) F-FDG and (11) C-PIB-PET for prediction of short-term conversion to Alzheimer's disease in subjects with mild cognitive impairment. *Int J Clin Pract* 2012;66:185–198.
27. Herholz K, Salmon E, Perani D, et al. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage* 2002;17:302–316.
28. Chen K, Ayutyanont N, Langbaum JB, et al. Characterizing Alzheimer's disease using a hypometabolic convergence index. *Neuroimage* 2011;56:52–60.
29. Fischl B, Dale A. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci USA* 2000;97:11050–11055.
30. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 2002;33:341–355.
31. Morra JH, Tu Z, Apostolova LG, et al. Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. *Neuroimage* 2008;43:59–68.
32. Chupin M, Mukuna-Bantumbakulu AR, Hasboun D, et al. Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: method and validation on controls and patients with Alzheimer's disease. *Neuroimage* 2007;34:996–1019.
33. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;271:703–707.
34. Albert J. *Bayesian Computation with R*. New York: Springer; 2009.
35. Jack CR Jr, Barkhof F, Bernstein MA, et al. Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criteria for Alzheimer's disease. *Alzheimers Dement* 2011;7:474–485.e4.
36. Thurfjell L, Lotjonen J, Lundqvist R, et al. Combination of biomarkers: PET [18F]flutemetamol imaging and structural MRI in dementia and mild cognitive impairment. *Neurodegener Dis* 2012;10:246–249.
37. Morinaga A, Ono K, Ikeda T, et al. A comparison of the diagnostic sensitivity of MRI, CBF-SPECT, FDG-PET and cerebrospinal fluid biomarkers for detecting Alzheimer's disease in a memory clinic. *Dement Geriatr Cogn Disord* 2010;30:285–292.
38. Visser PJ, Scheltens P, Verhey FR, et al. Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment. *J Neurol* 1999;246:477–485.
39. Fritzsche KH, Stieltjes B, Schlindwein S, van Bruggen T, Essig M, Meinzer HP. Automated MR morphometry to predict Alzheimer's disease in mild cognitive impairment. *Int J Comput Assist Radiol Surg* 2010;5:623–632.
40. Cuingnet R, Gerardin E, Tessieras J, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* 2011;56:766–781.
41. Ewers M, Walsh C, Trojanowski JQ, et al. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol Aging* 2012;33:1203–1214.
42. Bhadra D, Daniels MJ, Kim S, Ghosh M, Mukherjee B. A bayesian semiparametric approach for incorporating longitudinal information on exposure history for inference in case-control studies. *Biometrics* 2012;68:361–370.
43. Roberson ED, Mucke L. 100 years and counting: prospects for defeating Alzheimer's disease. *Science* 2006;314:781–784.
44. Frisoni GB, Redolfi A, Manset D, Rousseau ME, Toga A, Evans AC. Virtual imaging laboratories for marker discovery in neurodegenerative diseases. *Nat Rev Neurol* 2011;7:429–438.
45. Mosconi L, Tsui WH, De Santi S, et al. Reduced hippocampal metabolism in MCI and AD: automated FDG-PET image analysis. *Neurology* 2005;64:1860–1867.
46. Mattsson N, Rosen E, Hansson O, et al. Age and diagnostic performance of Alzheimer disease CSF biomarkers. *Neurology* 2012;78:468–476.