ELSEVIER

Alzheimer's & Dementia 7 (2011) 474–485

Alzheimer's & Dementia

Perspectives

Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease

Clifford R. Jack, Jr.,[a,*], Frederik Barkhof[b], Matt A. Bernstein[a], Marc Cantillon[c], Patricia E. Cole[d], Charles DeCarli[e], Bruno Dubois[f], Simon Duchesne[g,h], Nick C. Fox[i], Giovanni B. Frisoni[j], Harald Hampel[k], Derek L. G. Hill[l,m], Keith Johnson[n,o], Jean-François Mangin[p], Philip Scheltens[q], Adam J. Schwarz[r], Reisa Sperling[s], Joyce Suhy[t], Paul M. Thompson[u], Michael Weiner[v,w], Norman L. Foster[x]

[a]*Department of Radiology, Mayo Clinic, Rochester MN, USA*
[b]*Department of Radiology and Image Analysis Center, VU University Medical Center, Amsterdam, NL, USA*
[c]*Coalition Against Major Diseases, Critical Path Institute, Rockville, MD, USA*
[d]*Imagepace, Cincinnati, OH, USA*
[e]*Department of Neurology and Center for Neuroscience, University of California–Davis, Sacramento, CA, USA*
[f]*Department of Neurology, Salpetriere Hospital, Pierre and Marie Curie University, Paris, France*
[g]*Departement de Radiologie, Centre de Recherche Université Laval Robert-Giffard, Quebec, QC, Canada*
[h]*Department of Radiology, Université Laval, Québec, QC, Canada*
[i]*Dementia Research Centre, UCL Institute of Neurology, London, United Kingdom*
[j]*Laboratory of Epidemiology and Neuroimaging, IRCCS San Giovanni di Dio-FBF, Brescia, Italy*
[k]*Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, Goethe University, Frankfurt, Germany*
[l]*Centre for Medical Image Computing, Department of Medical Physics and Bioengineering, University College, London, United Kingdom*
[m]*IXICO Ltd, The London Bioscience Innovation Centre, London, United Kingdom*
[n]*Department of Neurology, Center for Alzheimer Research and Treatment, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*
[o]*Division of Molecular Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA*
[p]*Neurospin, I2BM, CEA, Saclay, France*
[q]*Department of Neurology and Alzheimer Center, VU University Medical Center, Amsterdam, NL, USA*
[r]*Translational Medicine, Eli Lilly and Company, Indianapolis, IN, USA*
[s]*Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*
[t]*Division of Neurology, Synarc Inc., San Francisco, CA, USA*
[u]*Department of Neurology, Laboratory of Neuro Imaging, UCLA School of Medicine, Los Angeles, CA, USA*
[v]*Department of Veterans Affairs, Center for Imaging of Neurodegenerative Disease, VA Medical Center, San Francisco, CA, USA*
[w]*Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA*
[x]*Department of Neurology, Center for Alzheimer's Care, Imaging and Research, University of Utah, Salt Lake City, UT, USA*

Abstract

Background: The promise of Alzheimer's disease biomarkers has led to their incorporation in new diagnostic criteria and in therapeutic trials; however, significant barriers exist to widespread use. Chief among these is the lack of internationally accepted standards for quantitative metrics. Hippocampal volumetry is the most widely studied quantitative magnetic resonance imaging measure in Alzheimer's disease and thus represents the most rational target for an initial effort at standardization.

Methods and Results: The authors of this position paper propose a path toward this goal. The steps include the following: (1) Establish and empower an oversight board to manage and assess the effort, (2) adopt the standardized definition of anatomic hippocampal boundaries on magnetic resonance imaging arising from the European Alzheimer's Disease Centers–Alzheimer's Disease Neuroimaging

*Corresponding author. Tel.: 507-284-8548; Fax: 507-284-9778.
E-mail address: Jack.Clifford@mayo.edu

doi:10.1016/j.jalz.2011.04.007

Initiative hippocampal harmonization effort as a reference standard, (3) establish a scientifically appropriate, publicly available reference standard data set based on manual delineation of the hippocampus in an appropriate sample of subjects (Alzheimer's Disease Neuroimaging Initiative), and (4) define minimum technical and prognostic performance metrics for validation of new measurement techniques using the reference standard data set as a benchmark.

Conclusions: Although manual delineation of the hippocampus is the best available reference standard, practical application of hippocampal volumetry will require automated methods. Our intent was to establish a mechanism for credentialing automated software applications to achieve internationally recognized accuracy and prognostic performance standards that lead to the systematic evaluation and then widespread acceptance and use of hippocampal volumetry. The standardization and assay validation process outlined for hippocampal volumetry was envisioned as a template that could be applied to other imaging biomarkers.

Keywords: Alzheimer's disease; Biomarkers; Magnetic resonance imaging; Hippocampus; Biomarker standards

1. Introduction

A biomarker is a physiological, biochemical, or anatomic parameter that can be objectively measured as an indicator of normal biologic processes, pathological processes, or responses to a therapeutic intervention [1]. Biomarkers used in the Alzheimer's disease (AD) field include both imaging measures and biofluid analytes. Biofluid analytes in this context can refer to proteins in any biofluid, but cerebrospinal fluid (CSF) biomarkers are presently the most well developed [2]. The five most widely studied biomarkers in AD can be divided into the following two major categories: (1) *Biomarkers of cerebral Aβ-amyloid* accumulation—these are increased radiotracer retention on amyloid tracer-based positron emission tomography (PET) imaging and low CSF$A\beta$ (1-42), and (2) *biomarkers of neuronal degeneration or injury*—these are elevated CSF tau (both total and phosphorylated tau), decreased fluorodeoxyglucose uptake on PET in the temporoparietal cortex, and brain atrophy in the medial, basal, and lateral temporal lobes and the medial and lateral parietal cortices determined from structural magnetic resonance imaging (MRI) or computed tomography [3]. Three of these five major AD biomarkers are imaging measures, and imaging is the primary focus of this position paper. Biomarkers are increasingly important in AD in two contexts: clinical diagnosis/prognosis and therapeutic trials.

Criteria for the clinical diagnosis of AD were established in 1984 [4]. These criteria have been widely adopted, validated against neuropathological examination in many studies, and are still being used today. A consensus now exists, however, that diagnostic criteria for AD should be updated to reflect the scientific advances of the past quarter of a century. One of the most important of these advances is the development of biomarkers for AD. This recognition has inspired recent efforts on several fronts to revise diagnostic criteria for AD. The two most well known among these efforts are those of Dubois et al [5,6] and the National Institute on Aging (NIA)–Alzheimer's Association (AA) [7–10]. The NIA–AA commissioned three workgroups to revise diagnostic criteria. Each was assigned the task of defining or revising criteria for one of the three recognized phases of the disease: preclinical or asymptomatic AD, symptomatic predementia or mild cognitive impairment (MCI), and the AD dementia phase [7–10]. Biomarkers providing evidence of in situ AD pathophysiology are used in the revised definitions of AD in all three phases of the disease by the NIA–AA and have also been included in the criteria of Dubois et al [5,6].

The second major use for biomarkers of AD is in clinical trials, where biomarkers can be used for several distinct purposes. As an indicator of AD pathophysiological processes, AD biomarkers may be used for subject inclusion/exclusion—to ensure study subjects are appropriate for targeting of the therapeutic mechanism of action or as an enrichment strategy to improve efficiency of therapeutic trials [2,11]. Biomarkers also provide a biologically based measure of disease severity. They can be used as a covariate in outcome analyses and as safety measures. Finally, an important application of AD biomarkers in clinical trials is as outcome measures, in which an effect on the biomarker is sought as evidence of modification of the underlying pathological AD process [12–21]. However, because AD pathophysiology is increasingly being recognized to be very complex and multifaceted, effects of candidate drugs on some individual pathophysiological aspects of AD may not necessarily be of functional or cognitive relevance. Therefore, increasing efforts are being spent on developing biomarkers which could serve as surrogate endpoints in clinical trials, accurately predicting and reflecting clinically significant outcomes [2,22]. Biomarkers are more objective and reliable quantitative measures of AD pathophysiological processes than traditional cognitive and functional outcomes that are affected by subject motivation and extrinsic factors such as alertness, environmental stresses, and informant mood and distress.

The evaluation of the value of biomarkers is different for therapeutic trials than for clinical diagnosis, but the rationale and methods to standardize and validate the reliability of the measures are very similar. Moreover, if an imaging biomarker is used as an inclusion criterion for subjects participating in a clinical trial of a compound that subsequently

achieves regulatory approval, then it is possible, some would say likely, that regulators will require the same biomarker to be approved as a diagnostic to identify patients who are suitable for treatment. This would then require that the biomarker, imaging in our case, be easily implementable in clinical imaging facilities worldwide. Therefore, although requirements in terms of precision and sensitivity to pathology may vary, issues pertaining to standardization of an imaging biomarker for use in clinical trials and for clinical diagnostics are inextricably interwoven.

The potential value of quantitative imaging biomarkers for both clinical diagnosis and clinical trials is clear, but there are major barriers to widespread acceptance and implementation. The most substantive barriers have been the lack of standardized methods for (1) image acquisition, (2) extraction of quantitative information from images, and (3) linking quantitative metrics to internationally recognized performance criteria. These in turn have impeded the establishment of cut-points in the continuous range of quantitative values that can be used in diagnosis and evaluating change in clinical trials. Standardization of image acquisition for structural MRI and PET scans has been a major focus of the Alzheimer's Disease Neuroimaging Initiative (ADNI) project [23,24], and ADNI acquisition protocols have become the *de facto* standard for clinical trials and could be applied clinically. In contrast, little progress has been made in the standardization of techniques for quantitative image analysis, either in ADNI or in the field in general. This is particularly true for MRI, where the lack of standardization has led to publication of values that are highly disparate across the literature. For example, greater than twofold differences in hippocampal volume of cognitively normal elderly subjects have been reported from different centers [25]. This is unlikely to have a basis in biology and is almost certainly because of intercenter differences in the measurement tools and the anatomical protocols for delineating the hippocampus. Likewise, a strong methodological dependence is evident in published rates of hippocampal atrophy. Threefold differences in rates of hippocampal atrophy have been reported in elderly controls as well as wide variations in apparently similar cohorts of patients with AD [26]. For example, Du et al [27] reported annualized rates of hippocampal atrophy in healthy elderly controls (mean age: 77 years) of 0.8%/yr; Jack et al [28] reported rates of 1.4%/yr in controls aged 78 years; and Wang et al [29] reported rates of 2.3%/yr in subjects with a mean age of 73 years. This strong dependence on the method used and its specific implementation undermine the credibility of the results. Both newly proposed diagnostic criteria explicitly point out that extensive work on imaging biomarker standardization is needed before widespread adoption for diagnostic purposes.

2. Why hippocampal volume?

Qualification or general acceptance of the validity of a biomarker in clinical trials must rest on a well-established body of evidence beginning with widespread agreement that there is clinical significance to the result of the biomarker and that it can be measured with appropriate accuracy and reproducibility. Quantitative measurement of hippocampal volume fulfills these basic criteria. The advantages of hippocampal volume as a target for an initial standardization and assay validation exercise are as follows: (1) The hippocampus is an anatomically defined structure with boundaries that are visually definable in a properly acquired MRI scan. (2) The hippocampus is involved early and progressively with neuronal loss and neurofibrillary tangles, which is one of the primary hallmarks of AD pathology [30]. (3) A large imaging and pathology literature provides evidence that loss of hippocampal volume is significant in AD. Numerous studies have shown the association of hippocampal atrophy with neurodegenerative pathology at autopsy [31–36], with clinical diagnoses of AD or MCI [37–43], and with the severity of cognitive disorders and episodic memory deficits because of AD pathophysiology [44,45]. In addition, longitudinal measures of change in hippocampal volume predict future cognitive decline and correlate with contemporary indices of clinical decline [46,47], and quantitative measures of the hippocampus predict progression from MCI to AD [48–63]. (4) Fully automated software tools are now available that can measure hippocampal volume efficiently and reproducibly [21,37,58,64–71]. Visual rating [72–74], although convenient and currently used in some diagnostic settings, does not lend itself to detecting subtle size differences, lacks precision relative to quantitative methods, and does not take advantage of the power of current technology. Formal computer-aided manual tracing of the entire hippocampus was introduced more than two decades ago to aid in seizure lateralization [75]. Although manual hippocampal tracing has been effective for research studies in different diseases, and still serves as the best available reference standard measure of the hippocampus on MRI [76], it is time-consuming and requires highly trained operators. Thus, it is not feasible in routine clinical practice and, owing to its expensiveness, is impractical in clinical trials. Fully automated hippocampal volumetry using standardized methods would be a practical alternative to manual methods. Automated hippocampal volumetry has successfully enabled the discovery of novel genes associated with hippocampal volume in more than 7000 subjects scanned at multiple internationally distributed sites. This result supports the assertion that such methods can be efficiently and reproducibly applied on a worldwide scale [77]. Furthermore, software methods that use within-subject registration permit sensitive measures of volume change over time [51,78]. (5) Although more complex MRI measures of disease-related atrophy consisting of combinations of multiple regions of interest might have superior diagnostic properties as compared with hippocampal volume [79–84], the analysis of hippocampal volume is less complex than multiple regions of interest approaches; therefore, a reference standard is easier to generate. Specifically, the hippocampus can be delineated by hand,

but the disease signatures of more complex analytic methods are a result of training and machine learning methods that would present a further challenge to validate, and are likely to evolve over time.

Further supporting hippocampal volumetry as a target for initial AD imaging biomarker standardization and assay validation is the fact that clinical guidelines in many countries [85,86] dictate that all patients investigated for cognitive impairment should undergo structural brain imaging to exclude treatable causes such as tumors and hematoma. An MRI acquisition sequence that would permit quantitative analysis of hippocampal volume is easy to include in a routine clinical MRI examination, only lengthens the examination by a few minutes, and is currently considered to be an essential part of a clinically diagnostic imaging protocol at some centers. Moreover, a significant effort has already been expended to standardize acquisition parameters for the high resolution three-dimensional (3D) anatomical MRI sequence needed for quantitative volume measures across MRI vendors in the ADNI study [23]. The ADNI 3D T1 anatomical sequence used for volumetric measurements can be performed in a standardized manner in an overwhelming majority of imaging centers worldwide. Finally, there is an ongoing international initiative led by one of the co-authors (G.B.F.) to establish a reference standard in hand-drawn hippocampal volumes, which is the European Alzheimer's Disease Centers (EADC)–ADNI Hippocampal Harmonization Effort [87,88].

The issue of validating imaging biomarkers for AD has recently drawn the attention of nonprofit organizations, including the Radiological Society of North America (RSNA) and the Coalition Against Major Disease (CAMD). CAMD is a part of the Critical Path Institute, a nonprofit public private partnership dedicated to more efficient drug development. Qualification of hippocampal atrophy for use in clinical trial enrichment is being pursued by CAMD with the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA). At a meeting of the RSNA Quantitative Imaging Biomarkers consortium in September 2010, a workgroup was convened to address the issue of standardizing quantitative imaging of AD. Among the candidate imaging modalities discussed, measures of hippocampal volume on structural MRI were identified as the most widely used in the context of multicenter clinical trials and therefore, were the most obvious candidates for an initial (exemplar) effort to standardize quantitative imaging biomarkers. This position paper follows from the recommendations of this RSNA workgroup.

3. Biomarker development

In general terms, three separate steps are required for biomarker development: (1) *Assay validation* (also called technical or analytical performance validity) to show that, when following defined standardized procedures, the biomarker can be measured precisely and accurately compared with a reference standard [89], (2) *clinical validation* to establish that the biomarker has value for a specific intended task and context of use, and (3) *qualification* of the biomarker with the appropriate regulatory agencies based on widespread consensus that the biomarker is "fit for purpose" for a particular use. Each proposed task (e.g., diagnostic, prognostic, outcome) needs to be considered separately. Qualification of a biomarker for clinical trials may be a stepping stone to a qualification for its use as a clinical diagnostic. However, the use of a biomarker in clinical diagnosis is distinct from its use in therapeutic trials, and development may focus on one or the other first. The use of a biomarker in clinical trials is at the discretion of the trial sponsor, but mechanisms have been introduced by which regulatory bodies (e.g., the U.S. FDA Center for Drug Evaluation and Research; or the EMA) qualify biomarkers for use in clinical trials. The use of a biomarker for clinical diagnosis requires regulatory approval in the relevant jurisdiction (e.g., approval by FDA Center for Devices and Radiological Health in the United States; or European Conformity marking in Europe), and may separately also require approval from healthcare funders for reimbursement.

4. Steps to standardization and validation of hippocampal volumetry as a biomarker of AD

In the following text, we outline the steps of a proposed work plan that would lead to standardization of quantitative (automated or manual) hippocampal volumetry as a biomarker for AD in evaluative studies in the context of clinical trials and for diagnosis.

1. Establish an *Oversight Board* to manage the effort and empower this body with authority to make decisions necessary to assess the results as outlined later in the text. The *Oversight Board* should have the following attributes: (a) Include all necessary areas of expertise, (b) be unbiased, (c) represent academia as well as industry, and (d) be international. All potential conflicts of interest must be fully disclosed. Our recommendation is that this oversight board be linked to the AA.

2. Identify a *standardized definition* of anatomic hippocampal boundaries on MRI with the assistance of expert neuroanatomists for use as a reference standard. Anatomic boundary criteria should be acceptable to the international scientific community and consistent with use in all neuroscience disciplines. We recognize that for hippocampal volume measures to be widely used diagnostically in clinical practice and in clinical trials, automated techniques are essential. However, manual tracing of the hippocampus using a consensus-from-experts approach in accordance with a standardized definition provides the most effective reference standard to evaluate automated methods. Expert opinion is an accepted method to create a reference standard. This is preferable to the alternative, arbitrarily picking one automated method and anointing it as the reference standard, which would be problematic. Because most, if not all,

automated techniques rely on some *a priori* anatomical notion of hippocampal boundaries, such an arbitrary approach would not reflect a consensus from the scientific community as a whole and would not result in a reference standard with broad-based support from all stakeholders. Because an international effort is currently in place with precisely this aim, leveraging the work of the EADC–ADNI Hippocampal Harmonization Effort [87,88] is the most logical and practical approach. The *reference standard* recommended by the authors of this position paper is therefore the manual hippocampal tracing of ADNI subjects, which will be developed by the EADC–ADNI effort.

3. Establish a *reference standard data set* based on manual delineation of the hippocampus in accordance with the standardized definition. The *reference standard data set* should have the following attributes:

a. All subjects in the reference database must have given informed consent for public access under an ethics board-approved protocol. Compliance with relevant privacy legislation to the jurisdiction where the data were collected, and permission of a research ethics committee for use of the data should be obtained. In the United States, the relevant guidelines are those of the Health Insurance Portability and Accountability Act; however, other jurisdictions will have different regulations.
b. Access to the database must be straightforward, open, and readily available.
c. Appropriate subjects, in clinical characteristics and number, must be included in the reference database—in this case, elderly cognitively normal control, MCI, and AD subjects diagnosed according to internationally recognized diagnostic criteria.
d. MRI scans must have been acquired with a standardized protocol that is amenable to widespread use.
e. Appropriate clinical metadata must be linked to the MRI scans and readily available to users, that is, demographics, clinical diagnosis, basic neuropsychology, and longitudinal clinical course.

The subjects, 3D volume T1-weighted images, and clinical data of ADNI represent a data set that meets these criteria. The authors recommend that the EADC–ADNI harmonization traces or masks of the 1.5-T ADNI magnetization prepared rapid acquisition gradient echo data serve as the hippocampal volume *reference standard data set.*

4. Extend the reference standard data set to enable a thorough evaluation of technical aspects of MR acquisition on measurement performance. This includes the effects of MR vendor, receiver coil type, accelerated acquisition methods, and field strength. Although the EADC–ADNI harmonization plan focuses are on 1.5-T data, a significant portion of neuroimaging in the future will be performed at 3-T, with acquisition acceleration, and with increasingly complex coil arrays. The potential effects of these technical advances on measurement standardization should be investigated [90].

5. Split the complete sample of traced hippocampi into balanced training and test data sets for assessing the technical performance characteristics of new analysis methods. This would enable automated methods to be trained on a portion of the reference data and then test performance against an independent subset of the reference data. Careful attention to the composition of these subsets is important so that age, gender, or clinical variables are not inadvertently unbalanced.

6. Develop standards for reporting measurement units, including a standardized approach for normalization of raw hippocampal volume measures. This will include defining correct measures of head size through standardization of intracranial volume measures. In addition to disease severity, hippocampal volume is affected by other variables that are easily ascertained such as age, gender, and head size (taller people tend to have larger brains and thus larger intracranial volume) [91]. Experience indicates that normalization of raw hippocampal volumes for these descriptive or confound variables improves the performance of hippocampal volumetry in evaluation studies and thus, recommendations for standardized normalization procedures for adjusting raw hippocampal volumes (e.g., by head size, age, gender) in the reference data set will be necessary.

7. Define minimum *technical performance metrics* as benchmarks to judge new analysis methods [89]. At a minimum, these metrics should include the following:

a. Accuracy with respect to the manually traced reference standard data set. We note that automated techniques will likely not precisely match a manually traced reference standard. However, a straightforward mathematical transformation of the output, an accurate automated algorithm to match the reference standard, should be possible. Criteria would need to be set as to how close the automated method would have to match the manual tracing for it to be credentialed by the oversight board.
b. Test/retest precision. This would include not just numeric precision at the volume level, but also more exacting indices of area/pixel overlap such as Dice coefficients.
c. Compliance with regulatory requirements (good clinical practice, FDA 21 Code of Federal Regulations part 11, European Union Governing Medicinal Products Annex 11 on computerized systems) for any computer systems running these algorithms.

8. Define minimum *prognostic performance metrics* for new analysis methods based on benchmarks established from reference standard data set: We recommend metrics that predict conversion from MCI to AD within 24 months, progression of dementia severity at 24 months in patients with AD, and maintenance of normal cognition at 24 months in cognitively normal subjects (sensitivity, specificity, positive and negative predictive value, receiver operating characteristic [ROC] analysis). This will serve as further assay validation for new analysis methods.

Table 1
Descriptive characteristics

Characteristic	All	Stable MCI	AD converter
N	373	178	135
Age, years	75 (70, 80)	75 (71, 81)	75 (70, 80)
Female gender, number (%)	136 (36)	63 (35)	51 (38)
Education, years	16 (14, 18)	16 (14, 18)	16 (14, 18)
MMSE	27 (26, 28)	28 (26, 29)	27 (25, 28)
Hippocampal volume, cm^3			
Method A	6.3 (5.6, 7.1)	6.7 (6.0, 7.4)	6.0 (5.2, 6.6)
Method B	6.3 (5.6, 7.1)	6.6 (5.9, 7.2)	5.9 (5.1, 6.6)
Method C	6.9 (6.2, 7.5)	7.1 (6.5, 7.6)	6.6 (6.0, 7.2)

Abbreviations: MMSE, Mini Mental State Examination; MCI, mild cognitive impairment; AD, Alzheimer's disease.

NOTE. All values are reported as median (inter quartile range) unless otherwise noted. Stable/Converter is defined as progression to AD by 24 months.

9. Empower the oversight board to oversee credentialing of applications for analysis methods. Although the reference standard data set can be used to credential new manual tracers, its primary use is envisioned as a means of validating and credentialing automated hippocampal quantification methods for use in therapeutic trials and for new clinical diagnostic criteria. The board could also make context of use recommendations based on limitations identified during the evaluation of a particular method. For a potential hippocampal volume measurement application to be credentialed by the oversight board, it would have to meet established technical and prognostic performance benchmarks using the reference data set described previously.

Ideally, the work plan would follow the aforementioned timeline, where initial steps would focus on establishing the reference standard of manual hippocampus traces, generating a standardized approach to volume normalization and benchmark performance metrics. After the reference standard is established, then the focus likely would be on evaluation studies and qualifying the reference standard with the FDA and EMA for diagnostic, prognostic, and outcome use in clinical trials. Standardized acquisition of MRI scans suitable for hippocampal volumetry are already widely performed and support from the pharmaceutical industry is likely. Subsequently, we expect evaluation studies will be conducted to show the diagnostic value of hippocampal volumetry use outside the context of clinical trials. We wish to emphasize that the intent of this position paper is not to stifle existing alternative methods or innovative development of new methods, but rather to facilitate the development of widely available implementations of automated hippocampal volumetry methods, and to serve as a template for an initial effort that can then be used for other imaging biomarkers.

5. Illustration

As an example, illustrating the approach discussed previously, we identified 373 ADNI subjects diagnosed as MCI at baseline who qualified for an analysis of time to progression to AD. Of the 397 ADNI subjects diagnosed as MCI at baseline, 16 had no follow-up visits and eight failed quality control, leaving 373 for this analysis (Table 1). A list of the ADNI subject identification numbers used in the example MCI analyses is included in Supplementary Table 1 (http://mayoresearch.mayo.edu/mayo/research/jack_lab/upload/HippocampStd.doc). All subjects had hippocampal volume measured in three ways, labeled as methods A, B, and C. In this exercise, we considered method A to represent the reference standard data set, and assessed methods B and C in two ways: technical performance accuracy relative to the reference standard data set and prognostic performance in predicting conversion from MCI to AD at 2 years after baseline. Although the data presented later in the text are real, and not hypothetical, the specific methods are left undefined because we do not wish to have this position paper misconstrued as evidence that the authors endorse a particular method for credentialing.

Of the 373 patients, 166 progressed from MCI to AD during follow-up and eight progressed to non-AD dementia, based on clinical criteria. We also examined a subset of 313 subjects that either progressed to AD at or before the 24-month visit (n = 135) or had available follow-up through the 24-month visit without progressing to AD (n = 178), so as to evaluate differences in hippocampal volume for those that progressed at 24 months versus those that remain stable. Subjects who progressed to non-AD dementia at or before 24 months were excluded from this analysis.

Method B potentially meets two major criteria for credentialing—it is highly accurate in the groupwise and individual measurement of hippocampal volume relative to method A, as shown in the table and scatterplots, and it also has essentially identical performance in predicting conversion from MCI to AD (Fig. 1, Table 2). Method C has a similar prognostic performance in predicting conversion to AD as method A, as shown in the ROC analysis, but in its current form might not meet technical accuracy criteria relative to the reference standard data set. This is how we would envision the credentialing process would proceed for most automated applications, with the EADC–ADNI harmonization data set of manually traced hippocampi serving as the reference standard data set and the oversight committee setting predetermined minimal benchmark criteria to judge the performance of individual methods.

One important feature of the process for critically evaluating automated hippocampal segmentation algorithms is the failure rate. For a variety of reasons, usually related to poor scan quality, automated algorithms will fail to produce a plausible result in some proportion of cases in a study. Taken to the extreme, imagine, for example, a method that produced perfect predictive results in cases that underwent successful hippocampal segmentation, but the method failed 99% of the time. The method would score quite well on prognostic metrics, but would not be practical. A fair and objective approach is therefore needed to penalize automated segmentation algorithms that fail in an unacceptably high proportion of cases.

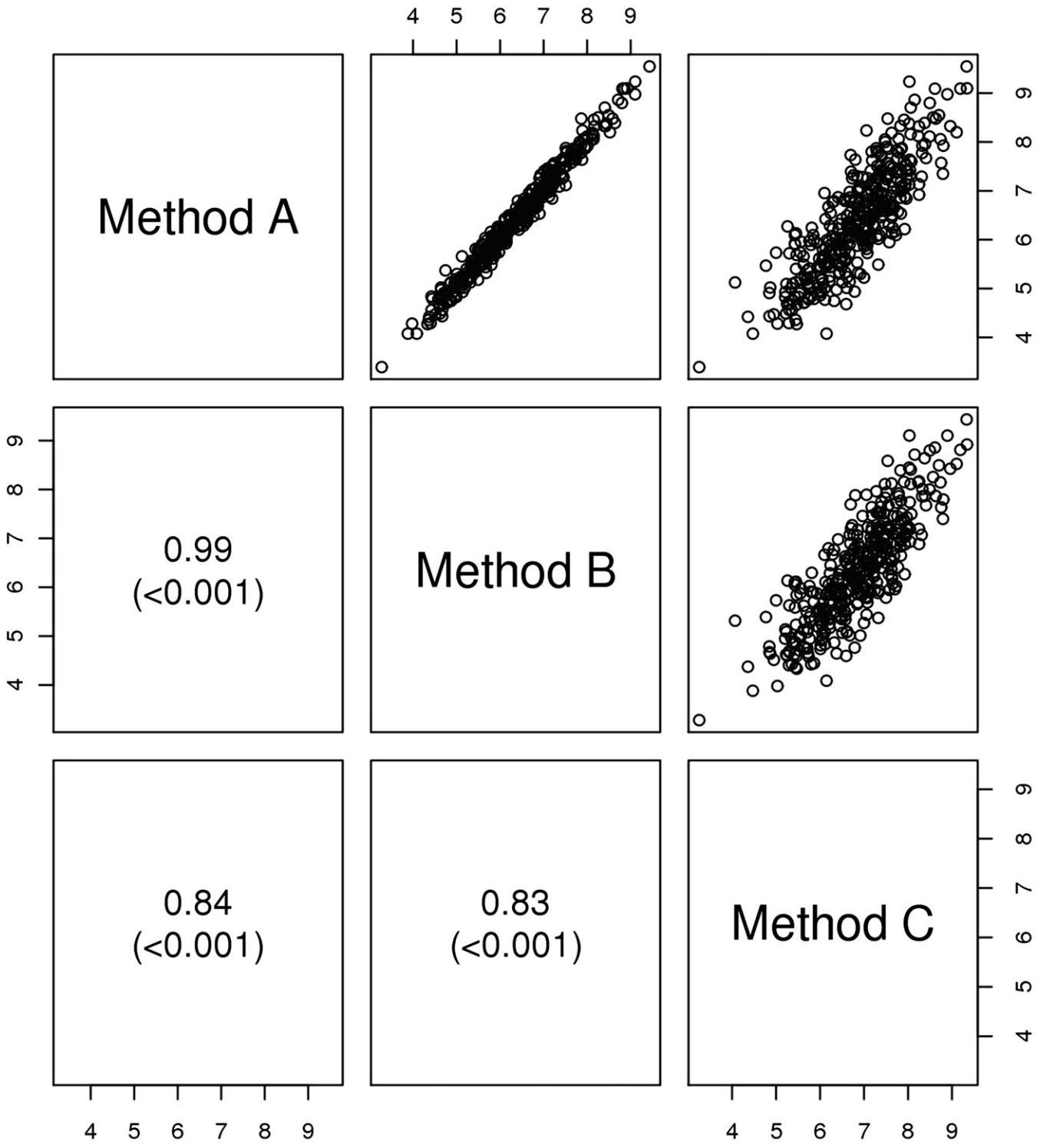

Fig. 1. Scatterplots of hippocampal volume (cm^3) by method. Spearman correlations and *P* values are shown for each pair.

6. Future efforts

1. Although a position paper is a first step, the objective of standardizing hippocampal volumetry as an AD biomarker will require active participation by stakeholders in academia and industry. The authors' objective is to see hippocampal volumetry evolve from its current state, a measure that is valid only in specific studies or a single institution, to a universally accepted biomarker with standardized units of measure. In some cases, this could simply involve having developers of automated measurement tools directly import

Table 2
ROC curves comparing prognostic performance of methods A, B, and C for progression from MCI to AD within 2 years

Area under the ROC curve for each hippocampal volume method predicting stable versus converter at 24 months	AUROC
Method A	0.675
Method B	0.678
Method C	0.625

Abbreviations: ROC, receiver operating characteristic; AUROC, area under the ROC curve.

the EADC–ADNI anatomic definition of the hippocampal boundaries into the atlas of the automated application.

2. Standardizing single time point hippocampal volume as an AD biomarker is the most logical and readily achievable initial goal; however, the authors recognize that other more complex topographic structural MRI measures might be more specific, or ultimately more powerful. The major difficulty here is identifying an appropriate reference standard if an anatomically based classifier does not conform to the boundaries of a classically defined anatomic structure as the hippocampus does.

3. Longitudinal change measures on structural MRI should be standardized using the approach outlined previously as a template. This could include an extension of the EADC–ADNI effort to include expert manual tracing of serial hippocampi to create a longitudinal reference standard data set using the same model as the single time point data set proposed in this position paper.

4. Fluorodeoxyglucose PET, amyloid PET imaging, and possibly other MRI modalities (e.g., resting state functional connectivity, diffusion tensor imaging, and arterial spin-labeled perfusion imaging) are also important imaging biomarkers for AD. Pursuing standardized quantitative metrics for these imaging modalities is a high priority. The efforts to standardize, validate, and evaluate quantitative measures in these modalities could roughly follow the same approach outlined previously for hippocampal volume.

5. For all imaging biomarkers, future efforts will need to focus on developing a quantitative score to allow the assessment of individual imaging biomarker measures against well-developed norms that incorporate other appropriate covariates, such as age, gender, and head size are for the hippocampus [91,92].

6. Ideally, diagnostic biomarkers should be evaluated against postmortem histopathological findings. It is well established that hippocampal atrophy, although being a feature of AD, is not specific for AD because it occurs in other conditions [32,93].

7. To optimize the use of biomarkers in new AD diagnostic criteria: future efforts will need to focus on establishing diagnostic cut-points in the continuous range of quantitative values to identify normal, abnormal, and indeterminate levels in individual subjects. For use in clinical practice, quantitative metrics will need to be developed and then tested in clinically typical and representative populations. Diagnostic biomarkers in AD should function analogously to those in other diseases where, for example, cut-points in the continuous range of blood pressure and fasting serum glucose are universally recognized as useful in aiding the diagnosis of hypertension and diabetes and standardized treatment protocols are based on these biomarker cut-points. For the purposes of diagnosis in typical clinical settings, cut-points should be derived from carefully characterized groups of subjects chosen in such a way that the results can be generalized to the overall population. For example, ADNI subjects were selected to represent a typical AD clinical trial, with specific inclusion/exclusion criteria. Thus, the results from ADNI are not generalizable to the overall population and are not optimal to generate normative data for general diagnostic purposes. Selecting meaningful diagnostic cut-points is complicated by the fact the many cognitively normal elderly subjects harbor significant AD pathology. Thus, the definition of normal is not straightforward. Consensus guidelines have been established for evaluating and reporting the clinical utility of diagnostic biomarkers and should be followed in studies using the results of the assay validation steps described in this article. In clinical settings, the sensitivity of detecting AD should exceed 80% and specificity for distinguishing AD from other similar dementias also should exceed 80% [94]. Standardized reporting of results should follow standards for the reporting of diagnostic accuracy studies (STARD) criteria [95], and for clinical settings additional reporting criteria to demonstrate pragmatic utility are needed [96].

Acknowledgments

The authors thank Heather J. Wiste, Mayo Clinic for helping in data analysis.

C.R.J., Jr. serves as a consultant for Janssen, Lilly, General Electric (GE), Johnson and Johnson, Eisai, and Élan. He is an investigator in clinical trials sponsored by Pfizer, Allon and Baxter, Inc. He receives research funding from the NIA (R01-AG11378 [PI], P50-AG16574 [Co-I], R21-AG38736 [Co-I], and U01 AG024904-01 [Co-I]), and the Alexander Family Alzheimer's Disease Research Professorship of the Mayo Foundation. He owns stock in Johnson and Johnson. F.B. has received compensation for consulting from Roche, Lundbeck, Janssen Alz Immunotherapy, and GE Medical Systems; M.A.B. receives research contract funding from Pfizer Inc., and Baxter Allon; M.C. is an employee of Coalition Against Major Diseases. P.E.C. reports no conflicts; C.D.C. reports no conflicts; B.D. has served on the advisory boards of GE healthcare and Eisai/Pfizer; S.D. reports no conflicts; N.C.F. served on the scientific advisory boards of Alzheimer's Research Forum, Alzheimer's Society and Alzheimer's Research Trust, and editorial boards of *Alzheimer's Disease and Associated Disorders; Neurodegenerative Diseases* and *BioMed Central—Alzheimer's Research and Therapy.* He holds a patent for QA

Box that may accrue revenue. In the last 5 years, his research group has received payment for consultancy or for conducting studies from Abbott Laboratories, Elan Pharmaceuticals, Eisai, Eli Lilly, GE Healthcare, Information extraction from imaging company (IXICO), Lundbeck, Pfizer Inc, Sanofi-Aventis, and Wyeth Pharmaceuticals. He receives research support from Medical Research Council (MRC) (G0801306 [PI], G0601846 [PI]), National Institutes of Health (NIH) (U01 AG024904 [Co-investigator{sub contract}]), Alzheimer Research Trust (ART/RF/2007/1 [PI]), National Institute for Health Research (NIHR) (Senior Investigator), and Engineering and Physical Sciences Research Council (EPSRC) (GR/S48844/01 [PI]); G.B.F. serves/has served on the advisory boards for Lilly, Bristol-Myers Squibb, Bayer, Lundbeck, Elan, Astra Zeneca, Pfizer, Taurx, and Wyeth. He is a member of the editorial boards of *Lancet Neurology*, *Aging Clinical & Experimental Research*, *Alzheimer's Disease & Associated Disorders, and Neurodegenerative Diseases*. He serves as the Imaging Selection Editor of *Neurobiology of Aging*. He has received grants from Wyeth Int'l, Lilly Int'l, Lundbeck Italia, and the AA; H.H. disclosed no conflicts; D.L.G.H. is an employee and stock holder in IXICO Ltd; K.J. reports no conflicts; J.-F.M. reports no conflicts; P.S. serves/has served on the advisory boards of Genentech, Novartis, Roche, Danone, Nutricia, Baxter, and Lundbeck. He has been a speaker at symposia organized by Lundbeck, Merz, Danone, Novartis, Roche, and Genentech. For all his activities, he receives no personal compensation.

He serves on the editorial board of *Alzheimer's Research & Therapy* and *Alzheimers Disease and Associated Disorders*, 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; A.J.S. is an employee and shareholder of Eli Lilly and Company; R.S. has been a consultant for Pfizer, Janssen, Elan, Bristol-Myers-Squibb, Bayer, and Avid (unpaid). She is a site investigator for clinical trials for Bristol-Myers-Squibb, Pfizer, Janssen, and Avid; J.S. is an employee of Synarc; P.M.T. reports no conflicts; M.W. serves/has served on the advisory boards of Elan/Wyeth, Novartis, Banner, Veterans Affairs Central Office, Lilly, Araclon and Insitut Catala de Neurociencies Aplicades, Biogen Idec, and Pfizer. He serves/has served as a consultant for Elan/Wyeth, Novartis, Forest, Ipsen, Daiichi Sankyo, Inc., Astra Zeneca, Araclon, Medivation/Pfizer, TauRx Therapeutics Ltd, Bayer Healthcare, Biogen Idec, Exonhit Therapeutics, Alzheimer's Association (SA), Servier, Synarc, and Pfizer. Funding for his travel has been provided by Elan/Wyeth, AA, Forest, University of California, Davis, Tel-Aviv University Medical School, Colloquium Paris, Ipsen, Wenner-Gren Foundations, Social Security Administration, Korean Neurological Association, National Institutes of Health, Washington University at St. Louis, Banner Alzheimer's Institute, Clinical Trials on Alzheimer's Disease (CTAD), Veterans Affairs Central Office, Beijing Institute of Geriatrics, Innogenetics, New York University, NeuroVigil, Inc., Centre Hospitalier Regional (CHRU)-Hospital 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, Pfizer, Paul Sabatier University, and Novartis. He serves on the editorial advisory boards of *Alzheimer's and Dementia* and *MRI*. He has received honoraria from American Academy of Neurology, Ipsen, NeuroVigil, Inc., and Insitut Catala de Neurociencies Aplicades. He receives research support from Merck and Avid. He owns stock in Synarc and Elan; N.L.F. disclosed no conflicts.

References

[1] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69:89–95.

[2] Hampel H, Frank R, Broich K, Teipel SJ, Katz RG, Hardy J, et al. Biomarkers for Alzheimer's disease: academic, industry and regulatory perspectives. Nat Rev Drug Discov 2010;9:560–74.

[3] Jack CR Jr, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010; 9:119–28.

[4] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984;34:939–44.

[5] Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 2007; 6:734–46.

[6] Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol 2010;9:1118–27.

[7] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging and Alzheimer's Association Workgroup. Alzheimers Dement 2011 (in press).

[8] Jack CR Jr, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carillo M, et al. Introduction to the recommendations from the National Institute on Aging and the Alzheimer Association workgroup on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 (in press).

[9] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging and the Alzheimer's Association Workgroup. Alzheimers Dement 2011 (in press).

[10] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging and the Alzheimer Association Workgroup. Alzheimers Dement 2011 (in press).

[11] McEvoy L, Hagler D, Holland D, Roddey J, Fennema-Notestine C, Salmon D, et al. Neuroimaging enrichment strategy for secondary prevention trials in Alzheimer disease. Alzheimers Dis Assoc Disord 2010;24:269–77.

[12] Fox NC, Black RS, Gilman S, Rossor MN, Griffith SG, Jenkins L, et al. Effects of Abeta immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. Neurology 2005;64:1563–72.
[13] Holland D, Brewer JB, Hagler DJ, Fenema-Notestine C, Dale AM, Weiner M, et al. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. Proc Natl Acad Sci USA 2009;106:20954–9.
[14] Jack CR Jr, Slomkowski M, Gracon S, Hoover TM, Felmlee JP, Stewart K, et al. MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. Neurology 2003;60:253–60.
[15] Schott JM, Frost C, Whitwell JL, Macmanus DG, Boyes RG, Rossor MN, et al. Combining short interval MRI in Alzheimer's disease: implications for therapeutic trials. J Neurol 2006;253:1147–53.
[16] Fox NC, Cousens S, Scahill R, Harvey RJ, Rossor MN. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: power calculations and estimates of sample size to detect treatment effects. Arch Neurol 2000;57:339–44.
[17] Leung KK, Clarkson MJ, Bartlett JW, Clegg S, Jack CR Jr, Weiner MW, et al. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. Neuroimage 2010;50:516–23.
[18] Hua X, Leow AD, Lee S, Klunder AD, Toga AW, Lepore N, et al. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. Neuroimage 2008;41:19–34.
[19] Hua X, Gutman B, Boyle C, Rajagopalan P, Leow AD, Yanovsky I, et al. Accurate measurement of brain changes in longitudinal MRI scans using tensor-based morphometry. Neuroimage 2011 (in press).
[20] Vemuri P, Wiste HJ, Weigand SD, Knopman DS, Trojanowski JQ, Shaw LM, et al. Serial MRI and CSF biomarkers in normal aging, MCI and AD. Neurology 2010;75:143–51.
[21] Wolz R, Heckemann RA, Aljabar P, Hajnal JV, Hammers A, Lotjonen J, et al. Measurement of hippocampal atrophy using 4D graph-cut segmentation: application to ADNI. Neuroimage 2010; 52:109–18.
[22] Hampel H, Wilcock G, Andrieu S, Aisen P, Blennow K, Broich K, et al. Biomarkers for Alzheimer's disease therapeutic trials. Prog Neurobiol (in press).
[23] Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27:685–91.
[24] Jagust WJ, Bandy D, Chen K, Foster NL, Landau SM, Mathis CA, et al. The Alzheimer's disease Neuroimaging Initiative positron emission tomography core. Alzheimers Dement 2010;6:221–9.
[25] Geuze E, Vermetten E, Bremner JD. MR-based in vivo hippocampal volumetrics: 1. Review of methodologies currently employed. Mol Psychiatry 2005;10:147–59.
[26] Barnes J, Bartlett JW, van de Pol LA, Loy CT, Scahill RI, Frost C, et al. A meta-analysis of hippocampal atrophy rates in Alzheimer's disease. Neurobiol Aging 2009;30:1711–23.
[27] Du AT, Schuff N, Kramer JH, Ganzer S, Zhu XP, Jagust WJ, et al. Higher atrophy rate of entorhinal cortex than hippocampus in AD. Neurology 2004;62:422–7.
[28] Jack CR Jr, Shiung MM, Gunter JL, O'Brien PC, Weigand SD, Knopman DS, et al. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. Neurology 2004; 62:591–600.
[29] Wang L, Swank JS, Glick IE, Gado MH, Miller MI, Morris JC, et al. Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. Neuroimage 2003; 20:667–82.
[30] Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 1991;82:239–59.
[31] Bobinski M, de Leon MJ, Wegiel J, Desanti S, Convit A, Saint Louis LA, et al. The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease. Neuroscience 2000;95:721–5.
[32] Jack CR Jr, Dickson DW, Parisi JE, Xu YC, Cha RH, O'Brien PC, et al. Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. Neurology 2002;58:750–7.
[33] Gosche KM, Mortimer JA, Smith CD, Markesbery WR, Snowdon DA. Hippocampal volume as an index of Alzheimer neuropathology: findings from the Nun Study. Neurology 2002;58:1476–82.
[34] Csernansky JG, Hamstra J, Wang L, McKeel D, Price JL, Gado M, et al. Correlations between antemortem hippocampal volume and postmortem neuropathology in AD subjects. Alzheimer Dis Assoc Disord 2004;18:190–5.
[35] Silbert LC, Quinn JF, Moore MM, Corbridge E, Ball MJ, Murdoch G, et al. Changes in premorbid brain volume predict Alzheimer's disease pathology. Neurology 2003;61:487–92.
[36] Zarow C, Vinters HV, Ellis WG, Weiner MW, Mungas D, White L, et al. Correlates of hippocampal neuron number in Alzheimer's disease and ischemic vascular dementia. Ann Neurol 2005;57:896–903.
[37] Colliot O, Chetelat G, Chupin M, Desgranges B, Magnin B, Benali H, et al. Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. Radiology 2008;248:194–201.
[38] Jack CR Jr, Petersen RC, O'Brien PC, Tangalos EG. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology 1992;42:183–8.
[39] Killiany RJ, Gomez-Isla T, Moss M, Kikinis R, Sandor T, Jolesz F, et al. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. Ann Neurol 2000;47:430–9.
[40] McEvoy LK, Fennema-Notestine C, Roddey JC, Hagler DJ Jr, Holland D, Karow DS, et al. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology 2009;251:195–205.
[41] Desikan RS, Cabral HJ, Hess CP, Dillon WP, Glastonbury CM, Weiner MW, et al. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain 2009;132(Pt 8):2048–57.
[42] Gerardin E, Chetelat G, Chupin M, Cuingnet R, Desgranges B, Kim HS, et al. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage 2009;47:1476–86.
[43] Dickerson BC, Goncharova I, Sullivan MP, Forchetti C, Wilson RS, Bennett DA, et al. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. Neurobiol Aging 2001;22:747–54.
[44] McDonald CR, Gharapetian L, McEvoy LK, Fennema-Notestine C, Hagler DJ Jr, Holland D, et al. Relationship between regional atrophy rates and cognitive decline in mild cognitive impairment. Neurobiol Aging (in press).
[45] Sarazin M, Chauvire V, Gerardin E, Colliot O, Kinkingnehun S, de Souza LC, et al. The amnestic syndrome of hippocampal type in Alzheimer's disease: an MRI study. J Alzheimers Dis 2010;22:285–94.
[46] Fox NC, Warrington EK, Freeborough PA, Hartikainen P, Kennedy AM, Stevens JM, et al. Presymptomatic hippocampal atrophy in Alzheimer's disease: a longitudinal MRI study. Brain 1996; 119(Pt 6):2001–7.
[47] Jack CR Jr, Petersen RC, Xu Y, O'Brien PC, Smith GE, Ivnik RJ, et al. Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. Neurology 2000;55:484–9.
[48] Jack CR Jr, Petersen RC, Xu YC, O'Brien PC, Smith GE, Ivnik RJ, et al. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. Neurology 1999;52:1397–403.
[49] Visser PJ, Scheltens P, Verhey FR, Schmand B, Launer LJ, Jolles J, et al. Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment. J Neurol 1999;246:477–85.
[50] Risacher SL, Saykin AJ, West JD, Shen L, Firpi HA, McDonald BC. Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. Curr Alzheimer Res 2009;6:347–61.

[51] Henneman WJ, Sluimer JD, Barnes J, van der Flier WM, Sluimer IC, Fox NC, et al. Hippocampal atrophy rates in Alzheimer disease: added value over whole brain volume measures. Neurology 2009;72:999–1007.
[52] Brys M, Glodzik L, Mosconi L, Switalski R, De Santi S, Pirraglia E, et al. Magnetic resonance imaging improves cerebrospinal fluid biomarkers in the early detection of Alzheimer's disease. J Alzheimers Dis 2009;16:351–62.
[53] Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, et al. Comparing predictors of conversion and decline in mild cognitive impairment. Neurology 2010;75:230–8.
[54] Kantarci K, Weigand SD, Przybelski SA, Shiung MM, Whitwell JL, Negash S, et al. Risk of dementia in MCI: combined effect of cerebrovascular disease, volumetric MRI, and 1H MRS. Neurology 2009; 72:1519–25.
[55] Karas G, Sluimer J, Goekoop R, van der Flier W, Rombouts SA, Vrenken H, et al. Amnestic mild cognitive impairment: structural MR imaging findings predictive of conversion to Alzheimer disease. Am J Neuroradiol 2008;29:944–9.
[56] Jack CR Jr, Wiste HJ, Vemuri P, Weigand SD, Senjem ML, Zeng G, et al. Brain beta-amyloid measure and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. Brain 2010;133:3336–48.
[57] Fleisher AS, Sun S, Taylor C, Ward CP, Gamst AC, Petersen RC, et al. Volumetric MRI vs clinical predictors of Alzheimer disease in mild cognitive impairment. Neurology 2008;70:191–9.
[58] Leung KK, Barnes J, Ridgway GR, Bartlett JW, Clarkson MJ, Macdonald K, et al. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. Neuroimage 2010;51:1345–59.
[59] Devanand DP, Pradhaban G, Liu X, Khandji A, de Santi S, Segal S, et al. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. Neurology 2007;68:828–36.
[60] Apostolova LG, Dutton RA, Dinov ID, Hayashi KM, Toga AW, Cummings JL, et al. Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. Arch Neurol 2006;63:693–9.
[61] Apostolova LG, Mosconi L, Thompson PM, Green AE, Hwang KS, Ramirez A, et al. Subregional hippocampal atrophy predicts Alzheimer's dementia in the cognitively normal. Neurobiol Aging 2010;31:1077–88.
[62] Chetelat G, Fouquet M, Kalpouzos G, Denghien I, De la Sayette V, Viader F, et al. Three-dimensional surface mapping of hippocampal atrophy progression from MCI to AD and over normal aging as assessed using voxel-based morphometry. Neuropsychologia 2008; 46:1721–31.
[63] Beckett LA, Harvey DJ, Gamst A, Donohue M, Kornak J, Zhang H, et al. The Alzheimer's Disease Neuroimaging Initiative: annual change in biomarkers and clinical outcomes. Alzheimers Dement 2010;6:257–64.
[64] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33:341–55.
[65] Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: part I. Segmentation and surface reconstruction. Neuroimage 1999;9: 179–94.
[66] Barnes J, Foster J, Boyes RG, Pepple T, Moore EK, Schott JM, et al. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. Neuroimage 2008;40:1655–71.
[67] Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, et al. Automated 3D mapping of hippocampal atrophy and its clinical correlates in 400 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. Hum Brain Mapp 2009;30:2766–88.
[68] Kim H, Besson P, Colliot O, Bernasconi A, Bernasconi N. Surface-based vector analysis using heat equation interpolation: a new approach to quantify local hippocampal volume changes. Med Image Comput Comput Assist Interv 2008;11(Pt 1):1008–15.
[69] Mouiha A, Duchesne S. Hippocampal atrophy rates in Alzheimer's disease: automated segmentation variability analysis. Neurosci Lett 2011;495:6–10.
[70] Chupin M, Hammers A, Liu RS, Colliot O, Burdett J, Bardinet E, et al. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. Neuroimage 2009; 46:749–61.
[71] Brewer JB, Magda S, Airriess C, Smith ME. Fully-automated quantification of regional brain volumes for improved detection of focal atrophy in Alzheimer disease. Am J Neuroradiol 2009; 30:578–80.
[72] Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein HC, Vermersch P, et al. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. J Neurol Neurosurg Psychiatry 1992; 55:967–72.
[73] de Leon MJ, Golomb J, George AE, Convit A, Tarshish CY, McRae T, et al. The radiologic prediction of Alzheimer disease: the atrophic hippocampal formation. Am J Neuroradiol 1993;14:897–906.
[74] DeCarli C, Frisoni GB, Clark CM, Harvey D, Grundman M, Petersen RC, et al. Qualitative estimates of medial temporal atrophy as a predictor of progression from mild cognitive impairment to dementia. Arch Neurol 2007;64:108–15.
[75] Jack CR Jr, Sharbrough FW, Twomey CK, Cascino GD, Hirschorn KA, Marsh WR, et al. Temporal lobe seizures: lateralization with MR volume measurements of the hippocampal formation. Radiology 1990;175:423–9.
[76] Teipel SJ, Ewers M, Wolf S, Jessen F, Kolsch H, Arlt S, et al. Multicentre variability of MRI-based medial temporal lobe volumetry in Alzheimer's disease. Psychiatry Res 2010;182:244–50.
[77] The ENIGMA Consortium, Genome-Wide Association Meta-Analysis of Hippocampal Volume via the ENIGMA Consortium. In: Organization for Human Brain Mapping Meeting, 2011, Quebec City, Canada.
[78] van de Pol LA, Barnes J, Scahill RI, Frost C, Lewis EB, Boyes RG, et al. Improved reliability of hippocampal atrophy rate measurement in mild cognitive impairment using fluid registration. Neuroimage 2007;34:1036–41.
[79] Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, et al. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage 2008;39:1186–97.
[80] Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol Aging 2008;29:514–23.
[81] Fan Y, Shen D, Davatzikos C. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. Med Image Comput Comput Assist Interv 2005; 8(Pt 1):1–8.
[82] Stonnington CM, Tan G, Kloppel S, Chu C, Draganski B, Jack CR Jr, et al. Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. Neuroimage 2008;39:1180–5.
[83] Kloppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008;131(Pt 3):681–9.
[84] Duchesne S, Caroli A, Geroldi C, Barillot C, Frisoni GB, Collins DL. MRI-based automated computer classification of probable AD versus normal controls. IEEE Trans Med Imaging 2008; 27:509–20.
[85] Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001; 56:1143–53.
[86] Hort J, O'Brien JT, Gainotti G, Pirttila T, Popescu BO, Rektorova I, et al. EFNS guidelines for the diagnosis and management of Alzheimer's disease. Eur J Neurol 2010;17:1236–48.

[87] Boccardi M, Ganzola R, Duchesne S, Redolfi A, Bartzokis G, Csernansky J, et al. Survey of segmentation protocols for manual hippocampal volumetry: preparatory phase for an EADC-ADNI harmonization protocol. In: ICAD - Alzheimer's Imaging Consortium, 2010, Honolulu, HI.

[88] Frisoni GB, Jack CR. Harmonization of magnetic resonance-based manual hippocampal segmentation: a mandatory step for wide clinical use. Alzheimers Dement 2011;7:171–4.

[89] Fox NC, Ridgway GR, Schott JM. Algorithms, atrophy and Alzheimer's disease: cautionary tales for clinical trials. Neuroimage (in press).

[90] Cover KS, van Schijndel RA, van Dijk BW, Redolfi A, Knol DL, Frisoni GB, et al. Assessing the reproducibility of the SienaX and Siena brain atrophy measures using the ADNI back-to-back MP-RAGE MRI scans. Psychiatry Res (in press).

[91] Barnes J, Ridgway GR, Bartlett J, Henley SM, Lehmann M, Hobbs N, et al. Head size, age and gender adjustment in MRI studies: a necessary nuisance? Neuroimage 2010;53:1244–55.

[92] Jack CR Jr, Petersen RC, Xu YC, Waring SC, O'Brien PC, Tangalos EG, et al. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology 1997;49:786–94.

[93] Jagust WJ, Zheng L, Harvey DJ, Mack WJ, Vinters HV, Weiner MW, et al. Neuropathological basis of magnetic resonance images in aging and dementia. Ann Neurol 2008;63:72–80.

[94] Reagan Working Group. Consensus report of the Working Group on: "Molecular and Biochemical Markers of Alzheimer's Disease". The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group. Neurobiol Aging 1998;19:109–16.

[95] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Fam Pract 2004; 21:4–10.

[96] Glasgow RE, Magid DJ, Beck A, Ritzwoller D, Estabrooks PA. Practical clinical trials for translating research to practice: design and measurement recommendations. Med Care 2005;43:551–7.

Supplementary Table 1
List of ADNI subject ID numbers used in the example MCI analyses

patientid	mridate
002_S_0729	7/17/2006
002_S_0782	8/14/2006
002_S_0954	10/10/2006
002_S_1070	11/28/2006
002_S_1155	12/14/2006
002_S_1268	2/14/2007
003_S_0908	9/12/2006
003_S_1057	12/4/2006
003_S_1074	12/4/2006
003_S_1122	12/6/2006
005_S_0222	2/21/2006
005_S_0324	3/30/2006
005_S_0448	5/4/2006
005_S_0546	6/15/2006
005_S_0572	6/20/2006
005_S_1224	1/23/2007
006_S_0675	8/31/2006
006_S_1130	11/30/2006
007_S_0041	10/21/2005
007_S_0101	12/20/2005
007_S_0128	1/16/2006
007_S_0249	3/2/2006
007_S_0293	3/14/2006
007_S_0344	3/31/2006
007_S_0414	5/22/2006
007_S_0698	7/7/2006
009_S_1030	11/1/2006
010_S_0161	1/19/2006
010_S_0422	11/30/2006
010_S_0904	12/7/2006
011_S_0168	2/10/2006
011_S_0241	3/10/2006
011_S_0326	3/20/2006
011_S_0362	3/28/2006
011_S_0856	9/15/2006
011_S_0861	9/27/2006
011_S_1080	11/22/2006
011_S_1282	2/9/2007
012_S_0634	6/16/2006
012_S_0917	12/1/2006
012_S_0932	9/20/2006
012_S_1033	11/16/2006
012_S_1165	12/28/2006
012_S_1175	1/5/2007
012_S_1292	3/1/2007
013_S_0240	3/20/2006
013_S_0325	4/19/2006
013_S_0860	9/21/2006
013_S_1120	11/22/2006
013_S_1186	1/29/2007
013_S_1275	2/22/2007
014_S_0169	2/7/2006
014_S_0557	5/31/2006
014_S_0563	7/5/2006
014_S_0658	7/25/2006
016_S_0354	5/5/2006
016_S_0702	7/24/2006
016_S_0769	8/2/2006
016_S_1028	11/2/2006
016_S_1092	12/11/2006
016_S_1117	12/1/2006
016_S_1121	12/6/2006
	(*Continued*)

Supplementary Table 1
List of ADNI subject ID numbers used in the example MCI analyses (*Continued*)

patientid	mridate
016_S_1138	12/28/2006
016_S_1326	3/1/2007
018_S_0057	11/17/2005
018_S_0080	12/29/2005
018_S_0087	12/22/2005
018_S_0103	1/5/2006
018_S_0142	1/19/2006
018_S_0155	2/23/2006
018_S_0406	4/20/2006
018_S_0450	5/4/2006
021_S_0141	1/23/2006
021_S_0178	2/10/2006
021_S_0231	2/28/2006
021_S_0273	3/14/2006
021_S_0276	3/17/2006
021_S_0424	4/20/2006
021_S_0626	6/29/2006
022_S_0004	9/22/2005
022_S_0044	11/3/2005
022_S_0544	5/17/2006
022_S_0750	8/7/2006
022_S_0924	9/27/2006
022_S_0961	10/20/2006
022_S_1097	11/30/2006
022_S_1351	3/16/2007
022_S_1366	3/28/2007
022_S_1394	5/29/2007
023_S_0042	10/31/2005
023_S_0126	2/8/2006
023_S_0217	2/21/2006
023_S_0331	3/23/2006
023_S_0376	4/3/2006
023_S_0388	4/10/2006
023_S_0604	6/2/2006
023_S_0625	6/23/2006
023_S_0855	9/5/2006
023_S_0887	9/20/2006
023_S_1046	11/6/2006
023_S_1104	11/15/2006
023_S_1126	12/5/2006
023_S_1247	1/31/2007
024_S_1393	3/13/2007
024_S_1400	5/8/2007
027_S_0116	1/27/2006
027_S_0179	2/24/2006
027_S_0256	3/21/2006
027_S_0307	4/6/2006
027_S_0408	5/10/2006
027_S_0417	5/12/2006
027_S_0461	6/2/2006
027_S_0485	5/8/2006
027_S_0644	6/16/2006
027_S_0835	9/11/2006
027_S_1045	11/3/2006
027_S_1213	1/19/2007
027_S_1277	2/9/2007
027_S_1387	2/26/2007
029_S_0878	9/15/2006
029_S_0914	12/18/2006
029_S_1038	11/21/2006
	(*Continued*)

Supplementary Table 1
List of ADNI subject ID numbers used in the example MCI analyses (*Continued*)

patientid	mridate
029_S_1073	11/21/2006
029_S_1215	1/19/2007
029_S_1218	1/23/2007
029_S_1318	2/17/2007
029_S_1384	3/28/2007
031_S_0294	3/16/2006
031_S_0351	4/18/2006
031_S_0568	5/22/2006
031_S_0821	8/30/2006
031_S_0830	9/13/2006
031_S_0867	9/20/2006
031_S_1066	11/10/2006
032_S_0187	2/13/2006
032_S_0214	2/20/2006
032_S_0718	7/12/2006
032_S_0978	10/16/2006
033_S_0511	6/1/2006
033_S_0513	5/18/2006
033_S_0514	5/18/2006
033_S_0567	5/17/2006
033_S_0723	7/14/2006
033_S_0725	7/20/2006
033_S_0739	7/19/2006
033_S_0906	9/25/2006
033_S_0922	9/26/2006
033_S_1116	11/21/2006
033_S_1279	2/12/2007
033_S_1284	2/12/2007
033_S_1309	2/8/2007
035_S_0033	11/22/2005
035_S_0204	2/14/2006
035_S_0292	3/22/2006
035_S_0997	11/29/2006
036_S_0656	7/7/2006
036_S_0673	7/21/2006
036_S_0748	8/10/2006
036_S_0869	10/30/2006
036_S_0945	10/25/2006
036_S_0976	12/12/2006
036_S_1135	12/27/2006
036_S_1240	2/13/2007
037_S_0150	2/1/2006
037_S_0182	2/14/2006
037_S_0377	5/2/2006
037_S_0501	5/23/2006
037_S_0539	6/15/2006
037_S_0552	5/24/2006
037_S_0566	6/13/2006
037_S_0588	11/28/2006
037_S_1078	11/29/2006
037_S_1225	1/24/2007
037_S_1421	8/27/2007
041_S_0282	4/19/2006
041_S_0314	3/28/2006
041_S_0446	5/2/2006
041_S_0549	6/13/2006
041_S_0598	6/16/2006
041_S_0679	7/20/2006
041_S_0721	9/14/2006
041_S_1010	12/12/2006
	(*Continued*)

Supplementary Table 1
List of ADNI subject ID numbers used in the example MCI analyses (*Continued*)

patientid	mridate
041_S_1260	2/1/2007
041_S_1411	7/16/2007
041_S_1412	6/25/2007
041_S_1418	7/27/2007
041_S_1420	9/24/2007
041_S_1423	8/3/2007
041_S_1425	8/6/2007
051_S_1040	10/31/2006
051_S_1072	11/24/2006
051_S_1131	12/15/2006
051_S_1331	4/13/2007
052_S_0671	7/5/2006
052_S_0952	10/19/2006
052_S_0989	11/14/2006
052_S_1054	11/28/2006
052_S_1168	12/8/2006
052_S_1346	3/13/2007
052_S_1352	3/6/2007
053_S_0389	4/20/2006
053_S_0507	5/15/2006
053_S_0621	6/12/2006
053_S_0919	10/16/2006
057_S_0464	5/17/2006
057_S_0839	9/20/2006
057_S_0941	10/11/2006
057_S_0957	10/18/2006
057_S_1007	10/25/2006
057_S_1217	1/31/2007
057_S_1265	1/31/2007
057_S_1269	1/31/2007
062_S_1182	1/17/2007
062_S_1294	2/15/2007
062_S_1299	2/20/2007
067_S_0038	11/10/2005
067_S_0045	11/16/2005
067_S_0077	12/21/2005
067_S_0098	1/30/2006
067_S_0176	4/3/2006
067_S_0243	3/9/2006
067_S_0284	3/17/2006
067_S_0290	3/28/2006
067_S_0336	5/4/2006
067_S_0607	6/27/2006
068_S_0401	4/25/2006
068_S_0442	5/9/2006
068_S_0476	5/11/2006
068_S_0478	5/16/2006
068_S_0802	8/25/2006
068_S_0872	10/24/2006
072_S_1211	3/9/2007
073_S_0518	9/11/2006
073_S_0746	11/30/2006
073_S_0909	10/2/2006
082_S_0832	8/30/2006
082_S_0928	9/29/2006
082_S_1119	12/7/2006
094_S_0434	4/20/2006
094_S_0531	5/16/2006
094_S_0921	10/2/2006
094_S_1015	10/30/2006
	(*Continued*)

Supplementary Table 1
List of ADNI subject ID numbers used in the example MCI analyses (*Continued*)

patientid	mridate
094_S_1188	12/28/2006
094_S_1293	3/12/2007
094_S_1314	2/26/2007
094_S_1398	5/3/2007
094_S_1417	7/16/2007
098_S_0160	1/28/2006
098_S_0269	3/4/2006
098_S_0667	6/24/2006
099_S_0051	11/15/2005
099_S_0054	11/16/2005
099_S_0060	12/7/2005
099_S_0111	1/18/2006
099_S_0291	3/9/2006
099_S_0551	5/18/2006
099_S_0880	10/5/2006
099_S_1034	11/2/2006
100_S_0006	11/15/2005
100_S_0296	4/3/2006
100_S_0892	10/16/2006
100_S_0930	10/3/2006
109_S_0950	10/25/2006
109_S_1114	12/27/2006
109_S_1183	1/3/2007
109_S_1343	3/20/2007
114_S_0378	4/4/2006
114_S_0410	4/18/2006
114_S_0458	5/9/2006
114_S_1103	11/29/2006
114_S_1106	11/21/2006
114_S_1118	12/8/2006
116_S_0361	4/27/2006
116_S_0649	7/24/2006
116_S_0752	8/16/2006
116_S_0834	9/29/2006
116_S_0890	1/22/2007
116_S_1243	2/7/2007
116_S_1271	2/21/2007
116_S_1315	3/8/2007
121_S_1322	3/2/2007
121_S_1350	3/9/2007
123_S_0108	2/1/2006
123_S_0390	4/11/2006
123_S_1300	3/8/2007
126_S_0708	7/17/2006
126_S_0709	7/26/2006
126_S_0865	11/6/2006
126_S_1077	12/6/2006
126_S_1187	1/17/2007
127_S_0112	1/13/2006
127_S_0393	4/12/2006
127_S_0394	5/17/2006
127_S_0397	5/15/2006
127_S_0925	10/30/2006
127_S_1140	12/14/2006
127_S_1419	7/23/2007
127_S_1427	8/20/2007
128_S_0135	1/26/2006
128_S_0138	1/25/2006
128_S_0188	2/6/2006
128_S_0200	2/13/2006

(*Continued*)

Supplementary Table 1
List of ADNI subject ID numbers used in the example MCI analyses (*Continued*)

patientid	mridate
128_S_0205	2/10/2006
128_S_0225	2/15/2006
128_S_0227	3/2/2006
128_S_0258	4/6/2006
128_S_0608	6/1/2006
128_S_0611	6/8/2006
128_S_0715	7/18/2006
128_S_0770	7/28/2006
128_S_0947	10/6/2006
128_S_1043	11/15/2006
128_S_1088	12/7/2006
128_S_1148	12/18/2006
128_S_1406	5/18/2007
128_S_1407	6/6/2007
128_S_1408	7/27/2007
129_S_1204	2/15/2007
129_S_1246	2/6/2007
130_S_0102	12/28/2005
130_S_0285	3/22/2006
130_S_0289	3/13/2006
130_S_0423	5/30/2006
130_S_0449	11/2/2006
130_S_0505	7/25/2006
130_S_0783	8/17/2006
131_S_0384	3/27/2006
131_S_0409	4/25/2006
131_S_1389	3/16/2007
132_S_0987	3/21/2007
133_S_0629	6/23/2006
133_S_0638	6/26/2006
133_S_0727	9/7/2006
133_S_0771	9/5/2006
133_S_0792	8/25/2006
133_S_0912	10/9/2006
133_S_0913	12/27/2006
133_S_1031	10/31/2006
136_S_0107	2/17/2006
136_S_0195	3/13/2006
136_S_0429	6/14/2006
136_S_0579	6/29/2006
136_S_0695	9/18/2006
136_S_0873	10/2/2006
136_S_0874	11/6/2006
136_S_1227	2/6/2007
137_S_0158	2/14/2006
137_S_0443	6/15/2006
137_S_0481	5/11/2006
137_S_0631	6/16/2006
137_S_0668	8/7/2006
137_S_0669	7/21/2006
137_S_0722	7/26/2006
137_S_0800	8/14/2006
137_S_0973	11/15/2006
137_S_0994	11/1/2006
137_S_1414	8/1/2007
137_S_1426	9/7/2007
141_S_0697	8/21/2006
141_S_0851	9/26/2006
141_S_0915	11/19/2006
141_S_0982	11/15/2006

(*Continued*)

Supplementary Table 1
List of ADNI subject ID numbers used in the example MCI analyses
(*Continued*)

patientid	mridate
141_S_1004	12/2/2006
141_S_1051	12/28/2006
141_S_1052	11/26/2006
141_S_1231	2/18/2007
141_S_1244	2/18/2007
141_S_1245	1/29/2007
141_S_1255	2/3/2007
141_S_1378	3/26/2007
941_S_1295	2/9/2007
941_S_1311	3/2/2007
941_S_1363	3/12/2007