



tomography [PET]), cerebral spinal fluid and blood biomarkers, and several clinical and neuropsychological measures acquired from healthy control subjects (normal), mild cognitive-impaired (MCI) subjects, and subjects with Alzheimer's disease (AD) followed over the course of 3 years (available at [www.loni.ucla.edu/ADNI](http://www.loni.ucla.edu/ADNI)) with up to an additional 6 years of data currently being acquired in the ADNI-GO and ADNI-2 projects [3]. The MRI data include MPAGE T<sub>1</sub>-weighted three-dimensional scans (or equivalent) acquired at regular (6-month or 12-month) intervals and intended for morphometric analysis [4], including volumetric measures of whole brain and regional structures, as well as cortical thickness and atrophy. These quantitative endpoints represent promising imaging biomarkers that are thought to be particularly sensitive to disease progression in the MCI and AD stages [3].

With numerous researchers working with the same data set, there is the potential for direct comparisons of the various endpoints of brain structures as well as the algorithms and preprocessing steps used to extract these structural measures. To ensure these comparisons are meaningful, it is desirable to define standardized data sets that multiple researchers can use for making methodological comparisons, thereby mitigating the risk that some of the observed differences in algorithm performance are an artifact of the use of different input data. Thus, to ensure meaningful side-by-side comparisons of structural MRI endpoints, the ADNI MRI Core proposes:

1. To define and make publicly available defined "standard analysis sets" of the structural magnetic resonance images comprising only image data that have passed quality control (QC) assessments conducted at the Aging and Dementia Imaging Research Laboratory at the Mayo Clinic (see [4] and <http://adni.loni.ucla.edu/research/protocols/mri-protocols/>).
2. To encourage any group publishing results using ADNI structural MRI data either to use one of these defined standard analysis sets of data or to justify the exclusion of any scans from their analysis and then to publish the actual data sets used.

In the following we expand on the motivation and the details of the proposal.

## 2. Motivation

Seven different research groups were funded to perform analyses of the ADNI-1 MRI data [3]. In addition, many other researchers have published analyses with ADNI-1 MRI data [5]. A PubMed search conducted on February 15, 2012, using the key terms "ADNI," "MRI," and "Volume" identified 46 matching articles. Although a few of these could be discounted as not directly involving the ADNI data, this large and growing set of articles shows the impact of this very rich data set in exploring various hypotheses and developing new analysis techniques. As

pharmaceutical and biotech companies try to apply this knowledge into their AD treatment trials they are faced with deciding which biomarkers and which analysis techniques are best suited to the needs of the trial. For example, there is great interest in determining which of the many reported techniques is most sensitive for measuring changes in brain or hippocampus volume. In addition to determining the most sensitive technique, there may be other metrics of interest for comparison, such as measurement bias. Unfortunately, every reported study to date on volume techniques based on the ADNI MRI data has used a different subset of data, confounding direct side-by-side comparisons—a key ADNI aim.

Although some of these differences are a result of availability of data at the time of publication, others are likely the result of variations in the robustness of a given technique or the imposition of additional quality control (QC) standards on the data set. The ADNI data go through an initial QC process; only data that pass the predefined criteria are released for further analysis. The QC process includes a comparison of image acquisition parameters in the Digital Imaging and Communications in Medicine (DICOM) header against the expected protocol, a visual check of the image quality by an experienced image analyst, and a quantitative check of the geometric accuracy of the scanner by analyzing data acquired with the ADNI phantom [4]. However, many algorithms may require additional QC steps to ensure successful processing or, alternatively, when the algorithm fails, the data are deemed to be unanalyzable and are not reported in the final analysis. This exclusion of subjects may thus obscure the real-world performance of a technique expected in a realistic clinical trial setting. For example, an algorithm that appears to work well but excludes half the data as being unsuitable for analysis would require a larger number of subjects if used as an endpoint in a trial than another technique that may, in testing, appear to be less effective but is able to run on the full data set with no exclusions. When reporting an interventional drug study, the final disposition of all subjects must be accounted for. *A priori* QC standards or reasons for dropout or unanalyzable endpoints need to be identified and summarized across all subjects not included in the analyses. This need for clear criteria for technical failure during analysis is described in the recent Food and Drug Administration draft guidance on standards for clinical trial imaging endpoints [6]. However, this type of rigor is seldom applied in studies aimed at developing new analysis techniques. To compare and contrast multiple techniques properly, it is important that each method is run on the same set of subjects and time points, and that the largest available data set be used so that the robustness of the techniques can be understood. We propose the higher level of rigor of drug studies can and should be adopted for all analysis studies using the ADNI MRI data, thereby permitting fairer evaluation and direct comparison among techniques.

### 3. Proposal

#### 3.1. ADNI Background

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.ucla.edu](http://adni.loni.ucla.edu)). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a \$60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years.” For up-to-date information, see “<http://www.adni-info.org>” [www.adni-info.org](http://www.adni-info.org).

#### 3.2. ADNI-1

To ensure consistency in the analysis of ADNI-1 MRI data, we have defined five standard analysis data sets (Table 1). Researchers are encouraged to use these data

Table 1  
ADNI-1 subjects available for standard analysis data sets at 1.5 T.

Screening visits		
Normal, n = 229	MCI, n = 401	AD, n = 188
<b>Complete year 1 visits (SC, M06, M12)</b>		
Normal, n = 195	MCI, n = 311	AD, n = 133
<b>Complete annual year 2 visits (SC, M12, M24)</b>		
Normal, n = 169	MCI, n = 234	AD, n = 101
<b>Complete 2-year visits (SC, M06, M12, M18*, M24)</b>		
Normal, n = 168	MCI, n = 212	AD, n = 99
<b>Complete visits (SC, M06, M12, M18*, M24, M36†)</b>		
Normal, n = 135	MCI, n = 148	AD, n = 99

Abbreviations: AD, Alzheimer's disease; M06 (M12, M18, M24, M36), month 6 (12, 18, 24, 36); MCI, mild cognitive impairment; SC, screening.

\*Month 18 imaging was conducted on subjects with MCI only.

†Month 36 imaging was only conducted on normal subjects and subjects with MCI.

sets and present results obtained using the most appropriate data set for their study. These five data sets differ based on the time points selected for inclusion. All subjects were included in the standardized analysis data set if the MRI of at least one of the two replicate T<sub>1</sub>-weighted volumetric sequences passed the QC conducted by the Mayo Clinic [4] for all the visits defined by the analysis set. The subjects also had to have all their scans performed on the same scanner because an analysis conducted on a subject scanned on different magnetic resonance equipment is likely to be technically inconsistent. During the 6 years of the ADNI-1 study, the scanners at some enrollment sites were replaced by the imaging department. This was the case for 42 ADNI subjects, whose data were removed from the multivisit standardized analysis sets because at least one visit was conducted on a different scanner than that used at their screening visit, typically because of the decommissioning or a major upgrade of the original scanner.

In ADNI-1, 818 subjects received screening MRI scans, met the study entry criteria, and were randomized. Of these 818 subjects, 197 subjects were assigned to receive 1.5-T MRI only, 419 to receive 1.5-T MRI and [18F]fluorodeoxyglucose (FDG)-PET, and 203 were assigned to receive both 1.5-T and 3.0-T MRI. By design, the sequence of postbaseline scanning time points differed by clinical diagnosis in that only MCI subjects received a month 18 (M18) scan whereas only normal subjects and MCI subjects received the month 36 (M36) scan. Table 1 summarizes the standardized data sets from the 818 subjects receiving 1.5-T scans.

The instrumental variability of 3.0-T MRI equipment at some of the sites at the start of the ADNI study in 2005 resulted in additional quality issues with the 3.0-T ADNI-1 MRI data. Specifically, some sites had older 3.0-T scanners (e.g., Siemens Allegra or GE VH3), which were limited to the use of a single-channel head coil. This resulted in a reduced signal-to-noise ratio and image contrast compared with their multichannel counterparts, given the constraints of the desired spatial coverage and spatial resolution (1.0 × 1.0 × 1.2 mm<sup>3</sup>) and maximal scan time of 9 to 10 minutes. These issues resulted in data that were generally of poor quality, which can be problematic for further processing and analysis. Because all modern 3.0-T scanners now have multichannel capabilities, those older, single-channel 3.0-T scanners are no longer representative of equipment readily available today for multicenter trials. Consequently, it was decided by the ADNI MRI Core that the 3.0-T images from sites with single-channel coils would be excluded from the standardized sets. Thus, of the 203 subjects who received a baseline 3.0-T MRI, only the 151 who received passing quality checks were included in the standardized data set for the baseline visit. Table 2 summarizes the standardized data sets from these 151 subjects who received 3.0-T scans.

It should be noted that the clinical diagnosis (normal, MCI, or AD) was evaluated at screening, and even though some subjects changed diagnosis (e.g., progressed from MCI to AD) at subsequent visits, they are still maintained in these lists under

**Table 2**  
ADNI-1 subjects available for standard analysis data sets at 3.0 T.

<b>Baseline visits</b>		
Normal, n = 47	MCI, n = 71	AD, n = 33
<b>Complete year 1 visits (BL, M06, M12)</b>		
Normal, n = 39	MCI, n = 56	AD, n = 24
<b>Complete annual year 2 visits (BL, M12, M24)</b>		
Normal, n = 34	MCI, n = 37	AD, n = 18
<b>Complete 2 year visits (BL, M06, M12, M18*, M24)</b>		
Normal, n = 33	MCI, n = 35	AD, n = 18
<b>Complete visits (BL, M06, M12, M18*, M24, M36<sup>†</sup>)</b>		
Normal, n = 22	MCI, n = 20	AD, n = 18

Abbreviations: AD, Alzheimer's disease; BL, baseline; M06 (M12, M18, M24, M36), month 6 (12, 18, 24, 36); MCI, mild cognitive impairment; SC, screening.

\*Month 18 imaging was conducted on subjects with MCI only.

<sup>†</sup>Month 36 imaging was only conducted on normal subjects and subjects with MCI.

the original diagnosis. However, the change in diagnosis and when it occurred are available on the ADNI website.

ADNI-1 data are available with different levels of preprocessing to reduce known image nonidealities. Most publications to date have been written using data that have been through gradient nonlinearity and intensity inhomogeneity correction [7]. These files are labeled GW\_N3 in the ADNI database found on the LONI website (see <http://adni.loni.ucla.edu/>). In addition, images with phantom-based distortion correction [8] are available and have file names with the suffix “scaled” or “scaled2.” It is not the intent of creating standard data sets to limit options available to the research community. An investigator could eschew all the corrections and begin with DICOM image data as they were acquired from the scanner, or they could choose to use data in which phantom-based scaling has not been applied, instead relying on their own methods to correct for scanner calibration drift over time.

### 3.3. ADNI-GO and ADNI-2

The standard analysis data sets described here are considered to be “frozen” at this point for the ADNI-1 project because all normal and MCI subjects have completed their final 3-year visits and all AD subjects have completed their 2-year visits. However, the ADNI-GO and ADNI-2 projects are still ongoing and will be collecting data for several more years. The ADNI-2 exams are acquired exclusively at 3-T, and although its protocol differs substantially from that of ADNI-1 [2, 3], it continues to include an unaccelerated volumetric T<sub>1</sub>-weighted scan, with additions of an accelerated T<sub>1</sub>-weighted volumetric scan, a fluid attenuation inversion recovery (FLAIR), and a T<sub>2</sub>\*-weighted gradient echo. Arms of the ADNI-2 study also include arterial spin labeling (ASL) MRI on Siemens scanners, resting-state functional MRI (rs-fMRI) on Philips scanners, and diffusion tensor imaging (DTI) on GE scanners. To balance rigor and consistency with reporting, along with timely dissemination of results, regular updates

of the currently available “standard analysis sets” will be provided on an annual basis. Researchers reporting prior to the final subject visits for a given analysis set should use the latest officially defined set available at a given point and state explicitly the version of the full analysis set used (e.g., the available standard analysis set as of the third quarter of 2012). Details of the latest standard analysis sets of available ADNI-GO/ADNI-2 data can be found on the ADNI standardization website as described in the next section.

### 3.4. Accessing data

For each of the ADNI-1, and ADNI-GO/ADNI-2 standardized analysis sets described earlier and in Tables 1 and 2, a complete listing of the subject IDs, scan dates, and unique image series IDs can be found on the website <http://adni.loni.ucla.edu/research/mri-analysis/adni-standardized-data>. In addition, collections of the images included in the lists have been created for download through the Image Data Archive on LONI. Further directions for accessing these image collections can be found at the previously mentioned website.

### 3.5. Cross-validation studies

Some algorithms require cross-validation on data that are separate from the data used to develop the algorithm. This is usually done by dividing the data into separate testing and training sets, and sometimes a third validation set. The test/training split is highly dependent on the technique and may vary between assigning the majority of the data to the testing set and assigning the majority of the data to the training set. The ADNI Biostatistics Core has provided researchers with guidance on conducting cross-validation studies and a predefined full analysis set split of the 818 subjects in ADNI by 40% for training and 60% for testing, as well as for a 10-fold cross-validation, which can be found at <http://www.adni-info.org/Scientists/CrossValidation.aspx>. Assignments were blocked by diagnosis at screening, age (<76 years, >76 year), and arm (1.5-T only, 1.5-T + 3.0-T, 1.5-T + FDG-PET). Researchers are encouraged to use the proposed cross-validation methods and test set when appropriate, and if a different cross-validation split is required to publish the actual subject IDs used in each split.

### 3.6. Publication standards

Any researcher conducting analysis on the ADNI structural MRI data is strongly encouraged to use one of the standard analysis data sets. It is understood that any algorithm or technique will potentially fail on a subset of data as a result of poor image quality, secondary QC, or other factors. These analysis exclusions or failures should be noted explicitly because it is extremely important for the assessment of the robustness of a given technique and to determine realistic sample size calculations. If, for other reasons, the study

must use a different subset of the data, then researchers are again strongly encouraged to publish a list of subjects and time points used in the final analysis either as supplemental material on the publishing journal's website or their own publicly available website. This list of subjects should then be referenced in the journal publication. These steps will result in greater rigor in the reporting of structural MRI analyses based on ADNI data and will permit replication of results and side-by-side comparisons with other techniques. This documentation of subjects analyzed will also help to understand the differences between image analysis techniques.

A standard metric on which different algorithms are compared is the sample size needed to detect a hypothetical treatment effect—for example, a 25% reduction in the rate of atrophy in subjects with AD with standard power (e.g., 80%) and type 1 error metrics [9]. An algorithm that fails to finish on a large portion of scans may, on the surface, perform well on this sample size metric. Yet in reality, such an algorithm would perform poorly in a clinical trial because of the large number of exclusions required. Therefore, we encourage all investigators who publish sample size estimates to include failed scans in the sample size estimates by amplifying the required sample size similar to how one might do so to account for attrition. This is the only way that different algorithms can be compared head-to-head on an even footing.

To help promote the use of these formally defined analysis data sets, we are asking the ADNI publications committee to encourage adherence in submitted manuscripts, and are asking journal editors and reviewers to insist on the use of either an official standard analysis data set or the inclusion of a supplemental file detailing the actual analysis set used.

#### 4. Discussion and issues

There are a number of occasions when use of a standard analysis set may not be appropriate for study and publishing. For example, a number of exploratory studies or new-technique development studies will use a small subset of the data for proof of concept. In these cases, analysis of the full data set may be deferred because of the time or cost involved. Likewise, it may be desirable to use all available data regardless of missing visits, or to create alternative data sets. For example, a data set with all subjects who had two or more visits may be useful for a biomarker prediction study. In these exceptions, it would still be appropriate to identify publicly the subject IDs actually used in the study.

The concept of standard analysis data sets becomes more complicated when combining biomarkers across modalities or when critical clinical data are lacking for correlative studies. For example, a comparison of MRI volumes with amyloid PET scanning endpoints or cerebral spinal fluid biomarkers necessitates selecting the data set defined by the intersection of the combined biomarkers. To facili-

tate these cross-modality studies we are encouraging the other ADNI cores to publish standard data sets of their respective biomarkers. Researchers reporting on multiple biomarkers are encouraged to continue to use the principle of using all the available data, accounting explicitly for any excluded subjects or time points and reporting all subject IDs included in the final analysis.

A good example of a side-by-side comparison of multiple ADNI biomarkers can be found in Beckett et al. [9]. They compared FDG-PET and MRI biomarkers using nine different analysis approaches from different laboratories, but ultimately had to limit the analysis comparison to the 69 subjects that had complete analysis across all nine methods. The greatest limitation of the data was a result of the inclusion of the FDG-PET modality in the analysis, because only about half the subjects received an FDG-PET scan whereas all subjects received MRI. However, we also note that the five MRI techniques used in this study were reported on different numbers of subjects with no clear explanation offered to explain differences in subject selection. Although this does not necessarily undermine the conclusions drawn, the additional rigor obtained by using more complete and identical data sets for analysis would greatly strengthen the comparisons and conclusions that can be drawn from them.

Many factors, however, can influence results of analyses, such as the number of subjects needed to detect a 25% reduction in a parameter. The overarching goal of our approach is to reduce study-to-study variability resulting from the specific scans included in the analyses. This approach does not address other concerns, such as the extent to which a particular outcome may be modifiable by a particular intervention. For example, one can imagine study 1 finds sample size of  $n_1$  people needed to detect a 25% reduction in parameter  $p_1$ , whereas study 2 finds sample size of  $n_2$  needed to detect a 25% reduction in  $p_2$ . If an intervention causes a much greater change in  $p_1$  than in  $p_2$ , then the technique of study 1 may be preferred, even if  $n_2$  is less than  $n_1$ . In essence, the initiative proposed here will address some of the technical details of comparing methods, but the appropriateness of any particular method as an outcome for a study will still require scientific reasoning and detailed understanding of biological equivalency. Conversely, methods such as factor analysis [10] can be used to determine whether differing measures (e.g., FDG, MRI, diffusion tensor imaging) have shared or unique explanatory power related to rates of biological change in AD. Further research, therefore, should consider the concept of biological equivalency when comparing various methods.

Another issue to consider is that there is a wealth of structural MRI ADNI studies published previously with variable subsets of the data. It is not the intent of this proposal to request authors to reanalyze those studies retrospectively. Although we do encourage these researchers to make available the list of subject IDs used, the aim of this proposal is to encourage adherence for future publications.

It should be noted that the concepts of presenting the analyses with respect to a standard data set and providing full disclosure of data that cannot be analyzed are not unique to ADNI MRI volumetric analysis. These principles should be applied to the analysis of any of the other ADNI data or any other scientific study. We present the details for achieving these goals with the ADNI structural MRI data and hope that the concepts will propagate beyond this scope.

We realize that there is no way to rigorously enforce this set of proposals, and the success of this concept will depend on its general acceptance by the MRI AD research community. However, the advantages of using standard analysis sets include

- Greater rigor in reporting
- The ability to compare various techniques side-by-side
- The ability to evaluate robustness of a given technique
- The ability to replicate methods

These advantages will help promote the goals of ADNI in developing the tools needed to eventually develop effective treatments for AD.

## Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for

the National Institutes of Health ("<http://www.fnih.org>" [www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

## References

- [1] Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, et al. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement* 2005;1:55–66.
- [2] Weiner MW, Aisen PS, Jack CR Jr, Jagust WJ, Trojanowski JQ, Shaw L, et al. The Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimers Dement* 2010;6:202–11.
- [3] Jack CR Jr, Bernstein MA, Borowski BJ, Gunter JL, Fox NC, Thompson PM, et al. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement* 2010;6:212–20.
- [4] 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.
- [5] Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, et al. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimers Dement* 2012; 8:S1–68.
- [6] U.S. Department of Health and Human Services. Guidance for Industry Standards for Clinical Trial Imaging Endpoints. Draft Guidance. Rockville MD: U.S. Department of Health and Human Services; 2011. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM268555.pdf>. Accessed: August 2012.
- [7] Boyes RG, Gunter JL, Frost C, Janke AL, Yeatman T, Hill DL, et al. Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. *Neuroimage* 2008; 39:1752–62.
- [8] Clarkson MJ, Ourselin S, Nielsen C, Leung KK, Barnes J, Whitwell JL, et al. Comparison of phantom and registration scaling corrections using the ADNI cohort. *Neuroimage* 2009;47:1506–13.
- [9] 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.
- [10] Thompson B. Exploratory and confirmatory factor analysis: understanding concepts and applications. Washington, DC: American Psychological Association; 2004.