Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort

Frithjof Kruggel [a,*], Jessica Turner [b], L. Tugan Muftuler [c,d]
and The Alzheimer's Disease Neuroimaging Initiative [1]

[a] *Department of Biomedical Engineering, University of California, Irvine, USA*
[b] *Department of Psychiatry and Human Behavior, University of California, Irvine, USA*
[c] *Center for Functional Onco-imaging, University of California, Irvine, USA*
[d] *Department of Radiological Sciences, University of California, Irvine, USA*

ARTICLE INFO

ABSTRACT

Morphometry of brain structures based on magnetic resonance imaging (MRI) data has become an important tool in neurobiology. Recent multicenter studies in neurodegenerative diseases raised the issue of the precision of volumetric measures, and their dependence on the scanner properties and imaging protocol. A large dataset consisting of 1073 MRI examinations in 843 subjects, acquired on 90 scanners at 58 sites, is analyzed here. A comprehensive set of image quality and content measures is used to describe the influence of the scanner hardware and imaging protocol on the variability of morphometric measures. Scanners equipped with array coils show a remarkable advantage over conventional coils in terms of image quality measures. The signal- and contrast-to-noise ratio in similar systems is equal or slightly better at 1.5 T than 3.0 T, while the white/grey matter tissue contrast is generally better on high-field systems. Repeated MRI investigations on the same scanner were available in 41 subjects, on different scanners in 172 subjects. The retest reliability of repeated volumetric measures under the same conditions was found as sufficient to track changes in longitudinal examinations in individual subjects. Using different acquisition conditions in the same subject, however, the variance of volumetric measures was up to 10 times greater. Two likely factors explaining this finding are scanner-dependent geometrical inaccuracies and differences in the white/grey matter tissue contrast.

© 2009 Elsevier Inc. All rights reserved.

Introduction

The Alzheimer's Disease Neuroimaging Initiative (ADNI) (Hua et al., 2008a,b; Jack et al., 2008; Mueller et al., 2005) is conducting a large-scale, multicenter, longitudinal study to collect demographic, cognitive, neuroimaging and genetic data about the progress of AD and possible conversion of individuals with mild cognitive impairment (MCI), a transitional state between normal aging and dementia that carries a 4- to 6-fold increased risk, relative to the general population. Magnetic resonance imaging (MRI) and image analysis methods can track brain atrophy at multiple time-points, and have revealed fine-scale anatomical changes associated with cognitive decline, e.g., (Fox et al., 2000; Jack et al., 2005). For recent neurobiological results of the ADNI study, refer to Hua et al. (2008a,b) and Leow et al. (2009).

The variance in brain volume found in a healthy population is much larger than disease-related changes. Compared to the gender-related difference in brain volume of 8.9%, the difference in brain volume between AD patients and healthy controls is 2.2%. The age-related brain loss of 0.17%/year is even a magnitude smaller (Kruggel, 2006). Thus, it is important to investigate the precision of MR imaging. Most previous studies in anatomical imaging of patient groups and healthy subjects were conducted at a single site, and comparing results across studies revealed puzzling unexplained differences, e.g., in brain volume of more than 10% or in grey/white matter volume ratio between 1.0 and 1.5 (Kruggel, 2006).

While multicenter studies can provide additional information over single center studies due to an increased statistical power, similar acquisition protocols must be used to avoid possible systematic differences between sites (Schnack et al., 2005). To demonstrate the feasibility of multicenter studies, several groups analyzed the influence of scanning protocols on morphometric results, however, focusing on global measures (Ewers et al., 2005), relatively small

* Corresponding author. 204 Rockwell Engineering Center, University of California, Irvine, CA 92697-2755, USA. Fax: +1 949 824 1727.
E-mail address: fkruggel@uci.edu (F. Kruggel).

[1] Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Citation.shtml).

subject groups (Han et al., 2006; Schnack et al., 2005), or specific imaging aspects (Jovicich et al., 2006, 2009; Shuter et al., 2008; Mortamet et al., 2009) only. The ADNI group diligently defined an optimized mandatory MPRAGE imaging protocol across all sites (Jack et al., 2008), and included repeated examinations of the same subject on the same and different scanners.

In this study, we address the following questions: (1) to what extent do parameters of the imaging protocol (e.g., scanner device, head coil, field strength, repetition time, echo time, voxel size) influence image quality (e.g., signal- and contrast-to-noise ratio, mutual information of the histogram×gradient magnitude histogram)? (2) How much do imaging protocol and image quality parameters influence the precision of head compartment volumes and ratios? (3) Is it possible to trace longitudinal changes in single subjects, allowing individual risk assessment? Using a comprehensive set of measures capturing image and segmentation precision, we analyze the large ADNI database here (1073 MRI datasets of 843 subjects acquired on 90 scanners at 58 sites).

In the following section, we characterize the subject sample and the MR scanner devices of the ADNI study, list the parameters studied here, and discuss the processing and analysis methods. The next section is devoted to an in-depth discussion of the statistical investigation and aims to assemble a comprehensive view of factors influencing the precision of imaging results. Conclusions of this analysis are drawn in the final section.

Materials and methods

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). 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 Principle Investigator of this initiative is Michael W. Weiner, M.D., 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 www.adni-info.org.

Subject sample, imaging devices and paramaters

The ADNI dataset (Jack et al., 2008) studied here includes 1073 MRI datasets of 843 subjects: 357 females (age: 74.8 ± 6.9 years, body weight: 65.9 ± 12.1 kg; 113 healthy controls, 149 MCI patients, 95 AD patients) and 486 males (age: 75.6 ± 6.9 years, body weight: 81.3 ± 15.4 kg; 119 healthy controls, 264 MCI patients, 103 AD patients). For the clinical definition of Alzheimer's disease (AD) and mild cognitive impairment (MCI) in this study, refer to Mueller et al. (2005).

Datasets were acquired on 98 scanners at 58 participating sites. Devices from the following manufacturers were used: General Electric (GE: Chalfont St. Giles, UK), Philips Medical Systems (PMS: Best, Netherlands), Siemens (SIE: Erlangen, Germany). For the purpose of this study, scanner hardware (denoted as *sh*) is encoded as the combination of device, coil and field strength. Details of the 17 systems included here are compiled in Table 1. An MPRAGE imaging protocol was used on all scanners (Jack et al., 2008). An example protocol is given for a Philips Achieva 3T scanner, equipped with a SENSE head coil, MPRAGE sequence, field-of-view 250 mm, matrix 256×256, 170 sagittal slices, 1.20 mm slice thickness, repetition time (TR) 6.802 ms, echo time (TE) 3.158 ms, flip angle 8 degrees. Detailed lists of all protocol information is publicly available (http://www.loni.ucla.edu:/ADNI/Research/Cores/). For the number of examinations, demographic information and clinical status per scanner hardware, refer to Table 1.

Collection of study variables

The set of variables analyzed in this study and their source is compiled in Table 2. Data were extracted from DICOM image files, ADNI database annotations exported as XML files, reported or derived quality measures (*qc*) in the original datasets, and compartment volumes from image segmentation (*seg*).

Table 1
List of the 17 scanner types included in this study, specified by manufacturer, device name, head coil, and field strength.

Manufacturer	Device	Coil	B_0 [T]	Sites	Examinations			Clinical status			Age	Encoding
					$\sum$	♀	♂	Normal	MCI	AD		
General Electric	Excite	HD	1.5	4	50	9	41	13	23	14	73.3 ± 8.0	Excite HD 1.5T
General Electric	Excite	PA	1.5	19	276	117	159	78	130	68	75.3 ± 6.9	Excite PA 1.5T
General Electric	Excite	PA	3.0	5	20	13	7	10	6	4	77.0 ± 5.3	Excite PA 3.0T
General Electric	Genesis	HD	1.5	6	80	33	47	19	44	17	74.9 ± 6.6	Genesis HD 1.5T
General Electric	Genesis	HD	3.0	1	6	3	3	2	2	2	74.8 ± 6.5	Genesis HD 3.0T
General Electric	HDx	PA	1.5	4	23	11	13	2	13	8	73.6 ± 6.6	HDx PA 1.5T
General Electric	HDx	PA	3.0	1	3	0	3	2	1	0	74.3 ± 2.6	HDx PA 3.0T
Philips	Achieva	HD	1.5	2	25	8	17	8	9	8	72.5 ± 5.2	Achieva HD 1.5T
Philips	Achieva	PA	1.5	7	82	25	57	24	42	24	76.0 ± 6.6	Achieva PA 1.5T
Philips	Achieva	PA	3.0	9	48	21	27	16	22	10	73.5 ± 7.5	Achieva PA 3.0T
Siemens	Allegra	HD	2.89	3	29	16	13	10	14	5	75.6 ± 8.0	Allegra HD 3.0T
Siemens	Avanto	PA	1.49	6	66	28	38	22	19	15	74.8 ± 7.1	Avanto PA 1.5T
Siemens	Sonata	HD	1.49	4	60	28	32	15	28	17	76.4 ± 6.2	Sonata HD 1.5T
Siemens	Sonata	PA	1.49	3	47	15	32	11	28	8	75.9 ± 7.6	Sonata PA 1.5T
Siemens	Symphony	HD	1.49	2	36	21	15	9	19	8	74.3 ± 5.8	Symphony HD 1.5T
Siemens	Symphony	PA	1.49	9	129	71	58	38	63	28	76.7 ± 6.8	Symphony PA 1.5T
Siemens	Trio	PA	2.89	13	93	48	45	21	51	21	74.6 ± 7.0	Trio PA 3.0T
Totals				98	1073	467	606	300	524	249	75.2 ± 6.0	

A total of 1073 examinations on 98 scanners are detailed by sites in column 5, subject in columns 6–8, and clinical status in columns 9–11. MR coils types are abbreviated as HD for conventional (quadrature birdcage) head coils, and PA for 8-channel phased array head coils.

Table 2
Compilation of the variables, their meaning and origin in this study. Refer to the text for further explanation.

Variable	Meaning	Source	Relation	Unit/factor levels
age	Subject age	XML	Subject	Years
gender	Gender	XML	Subject	m/f
weight	Subject weight	XML	Subject	kg
grp	Clinical group	XML	Subject	Normal/MCI/AD
dc	Device/coil combination	DICOM	Device	–
fs	Field strength	DICOM	Device	T
tr	Repetition time	DICOM	Protocol	ms
te	Echo time	DICOM	Protocol	ms
vs	Voxel size	DICOM	Protocol	mm³
snr	Signal/noise ratio	qc	Protocol	–
cnr	Contrast/noise ratio	qc	Protocol	–
mi	Mutual information	qc	Protocol	–
wgc	WM/GM contrast ratio	seg	Study	–
csfv	CSF volume	seg	Study	ml
icv	intracranial volume	seg	Study	ml
gmv	GM volume	seg	Study	ml
wmv	WM volume	seg	Study	ml
brv	(GM + WM) volume	seg	Study	ml
gwr	gmv/wmv ratio	seg	Study	–
brr	brv/icv ratio	seg	Study	–

Imaging quality parameters were determined as follows: DICOM images were converted into BRIAN format (Kruggel and Lohmann, 1996). From the converted "raw" images, a gradient magnitude image was computed using central differences. A joint intensity×gradient magnitude histogram was computed in 256×256 bins, discarding 5% of the highest intensity and gradient voxels. The absolute noise level is determined from the first peak in the marginalized gradient magnitude histogram (Gudbjartsson and Patz, 1995). In the marginalized intensity histogram, the prominent intensity peaks (class 1: roughly corresponding to grey matter (GM), muscles, and connective tissue; class 2: white matter (WM), and fat) were determined using a Gaussian mixture model (Bishop, 1995). The absolute signal level is defined as the mean of the intensity distribution of classes 1 and 2; the absolute contrast is the difference of the mean intensity of classes 1 and 2. The signal-to-noise ratio (SNR) is determined as the quotient of the absolute signal and absolute noise, the contrast-to-noise ratio (CNR) likewise. The mutual information (MI) (Press et al., 2007) of the joint histogram is a general parameter describing image quality: in an ideal image with n intensity classes without noise, intensity inhomogeneities, partial volume effect and an ideal point spread function, this joint histogram would consist of n peaks. Any deviance from ideal conditions smooths out peaks, and decreases the (negative) mutual information. Thus, higher MI values correspond to a better image quality.

Datasets were aligned with the stereotaxic coordinate system and interpolated to an isotropic voxel size of 1 mm using fourth-order b-spline interpolation (Kruggel and von Cramon, 1999). The outer hulls of the brain were removed using a registration-based approach (Hentschel and Kruggel, 2004), yielding a mask of the intracranial compartment (IC). This compartment was segmented into three classes using the "Fuzzy and Noise Tolerant Adaptive Segmentation Method (FANTASM)" (Pham, 2001). This algorithm produces a soft segmentation while simultaneously adapting to intensity inhomogeneities in the image. Constraints on the gain field are imposed to ensure that the estimated field is smooth and slowly varying. A Tikhonov-Phillips regularization in a multigrid approach is used here. Optimal values for the smoothness parameters α and β were derived by maximizing MI in the output image, and were fixed for the whole sample. Robustness to noise is achieved by including a term in the objective function that regularizes the class membership value of a voxel based on the values in its neighborhood. On output, three probability images are obtained that correspond to the likelihood that a voxel contributes to compartments 0 (CSF), 1 (GM), and 2 (WM).

Compartment volumes were determined by integrating the voxelwise compartment probabilities over the IC domain. The WM/GM contrast ratio wgc is computed as the ratio of the average intensities of class 2 to class 1.

We used the statistics software R (The R Foundation for Statistical Computing, ISBN 3-900051-07-0) to evaluate data. Model selection was performed by eliminating variables with the least influence, based on the adjusted R^2 (linear models) or Akaike's Information Criterion (AIC, linear mixed effect models). Except where noted, we discuss only strong influences only that have error rates on the null hypothesis of $p<0.001$ and/or explain at least 1% of the total variance. Processing was performed on a 10-node cluster (2×AMD64, 2.4 GHz processor, Linux 2.6.25 operating system, 4 GB RAM per node). The image processing chain takes about 12 min of computation time per dataset.

Results

The results of these analyses in the following sections are organized around the three previously identified questions. Thus we explored the relationships between device-related and subject-related parameters first; then the relationship between the protocol parameters and the image quality measures and; and finally the most important question, the relationship between the segmented compartment volumes and the scanner hardware and protocol implementation. We refine our analysis by selecting two subgroups of the ADNI dataset: (1) subjects examined twice using the same conditions, and (2) subjects examined twice on different scanners. This allows determining the precision of brain compartment measures and the factors that impact precision.

Independence of subjects, devices, and imaging protocol

Ideally, subject-related variables (age, gen, grp, weight) should be independent of parameters of the imaging protocol (sh, tr, te, vs).

More males than females were included in the sample, especially in the MCI group. Age was not different between gender and clinical groups. Healthy controls were significantly heavier ($+4.4$ kg, $p=0.0002$) than MCI ($+2.2$ kg, n.s.) or AD patients. There is a loss of weight with age in males (-0.30 kg/y, $p=0.0073$) that is smaller and non-significant in females. Gender is not balanced across clinical groups, scanner types and participating sites. Due to this imbalance, there is an interaction between body weight, scanner hardware and study site.

Protocol parameters repetition time *tr*, echo time *te* and voxel size *vs* trivially depend on the scanner configuration *sh*. With the exception of weight, protocol parameters are independent of subject variables.

Impact of scanner hardware on image quality

Image quality was rated by the SNR, CNR, and the mutual information MI of the intensity×gradient magnitude histogram. Higher values in all quality measures correspond to better image quality. The influence of scanner hardware *sh* and protocol parameters protocol (*tr*, *te*, *vs*) on image quality (snr, cnr, mi) was examined. Because the MPRAGE protocol was used for all examinations, protocol parameters tr, te, and vs are highly correlated with the scanner type sh.

The SNR is foremost dependent on the scanner hardware sh (see Fig. 1) and alone explains 74% of the variance. The relative performance of the systems is ranked on the right. Ties are given if results are statistically not different (*t* test, unequal variance, significance level $p=0.05$). Array coils (PA) offer a significant advantage over conventional coils (HD). Comparing the SNR on similar devices operating at different field strengths, 1.5T systems are equal or better than 3.0T systems (e.g., Achieva PA, HDx PA, Excite PA). Overall examinations, SNR decreases with field strength (-28.5/T), echo time
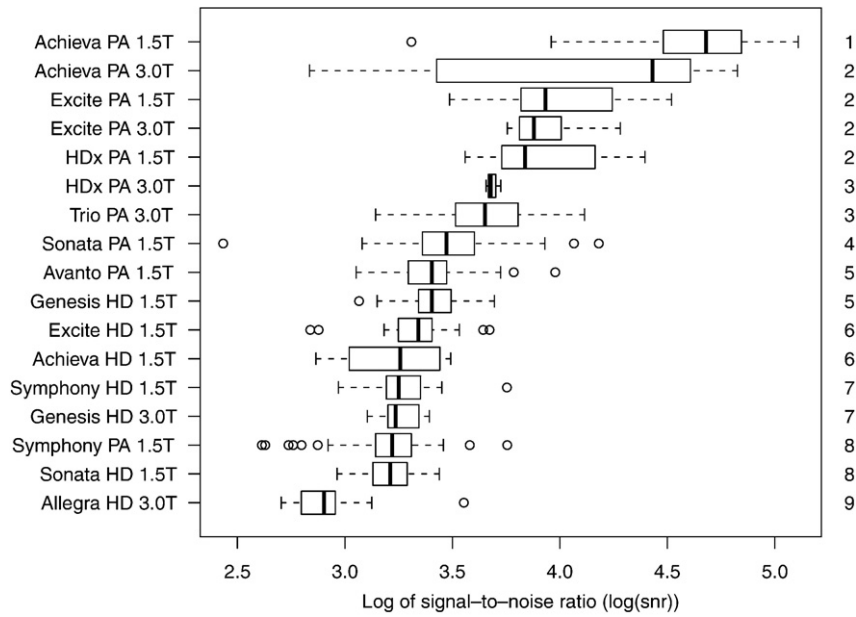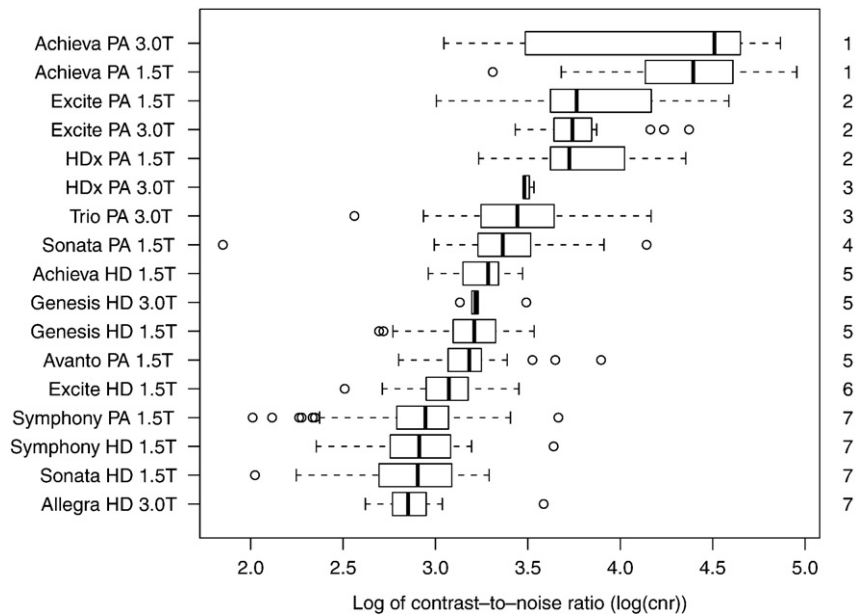
Fig. 1. Dependency of signal-to-noise ratio (SNR) on scanner hardware, sorted by median. Bars in this boxplot denote the median, boxes the 25%/75% quartiles, whiskers the minimum/maximum range. Rankings are shown on the right. Ties denote non-significant differences ($p = 0.05$). Refer to Table 1 for device parameters and a description of the subject sample.

(-45.2/ms), and increases with repetition time ($+11.3$/ms) and voxel size ($+29.8$/mm^3).

The CNR depends on the scanner type only, explaining 65% of the variance. Using array coils leads to a profound contrast increase (see Fig. 2) (e.g., Achieva PA vs. Achieva HD, Excite PA vs. Excite HD, Sonata PA vs. Sonata HD). Again, 1.5T systems are equal or better than 3.0T systems (e.g., Achieva PA, HDx PA, Excite PA). Over all examinations, CNR decreases with field strength (-0.47/T), echo time (-0.90/ms), and increases with repetition time ($+2.36$/ms) and voxel size ($+3.02$/mm^3). The marked differences in this ratio across systems are–to a large extent–due to differences in the absolute noise level, although significant differences in the WM/GM contrast ratio are found as well (see below).

The overall quality measure MI is compared across scanner hardware in Fig. 3, and again demonstrates the advantage of using array coils, especially on Philips Achieva systems. The scanner hardware explains 85% of the variance alone. Overall examinations, MI decreases with field strength (-0.024/T), echo time (-0.073/ms), and increases with repetition time ($+0.021$/s) and voxel size ($+0.057$/mm^3).

The white/grey matter contrast ratio *wgc* in segmented and inhomogeneity-corrected images across scanner hardware is compiled in Fig. 4). Here, a higher field strength offers a relative advantage on similar systems (e.g., Achieva PA, Genesis HD, HDx PA) while using array coils does not offer an improvement over conventional coils (e.g., Sonata 1.5T, Excite 1.5T, Achieva 1.5T, Symphony 1.5T). Over all

Fig. 2. Dependency of contrast-to-noise ratio (CNR) on scanner hardware, sorted by median. Bars in this boxplot denote the median, boxes the 25%/75% quartiles, whiskers the minimum/maximum range. Rankings are shown on the right. Ties denote non-significant differences ($p = 0.05$). Refer to Table 1 for device parameters and a description of the subject sample.
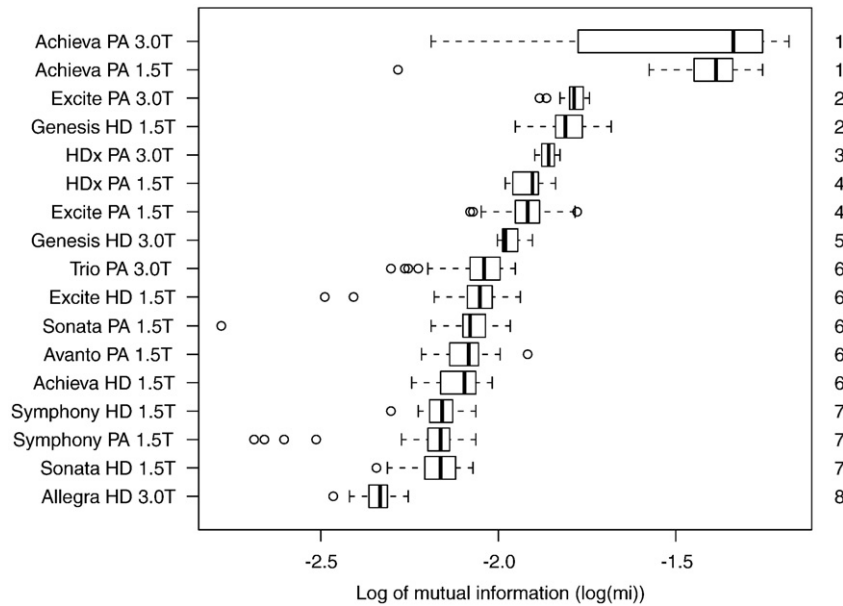
Fig. 3. Dependency of mutual information (MI) on scanner hardware, sorted by median. Bars in this boxplot denote the median, boxes the 25%/75% quartiles, whiskers the minimum/maximum range. Rankings are shown on the right. Ties denote non-significant differences ($p = 0.05$). Refer to Table 1 for device parameters and a description of the subject sample.

examinations, the WM/GM contrast ratio increases with field strength ($+0.161$/T), echo time ($+0.191$/ms), and decreases with repetition time ($−0.050$/ms) and voxel size ($−0.157$/mm^3). The Philips Achieva 3.0T system offers an exceptional contrast.

In summary, using array coils leads to a remarkable improvement in image quality as measured by SNR, CNR, and MI. Quality measures SNR and CNR in similar systems are equal or slightly better at 1.5T than 3.0T, while the MI and the WM/GM contrast ratio are generally better on high-field systems. The Philips Achieva 3.0T system was ranked best over all measures.

Disregarding the confounded variable *weight*, subject-related variables (age, gender, clinical group) do not influence image quality parameters.

Impact of imaging protocol on compartment volume precision

Ideally, image content (i.e., brain compartment volumes) should be independent of the scanner hardware and protocol implementation. Compartment volumes icv, brv, gmv, wmv and the ratios brr (brain/intracranial volume) and gwr (GM/WM matter ratio) were tested against hardware-, protocol- and subject-related parameters. Results are compiled in Table 3. The most parsimonous models explained between 44 and 68% of the total variance, of which about 22% correspond to subject-related variables, the rest is scanner- and protocol-related.

The intracranial volume depends only on gender and weight—an effect that is explained by their correlation with body volume. The
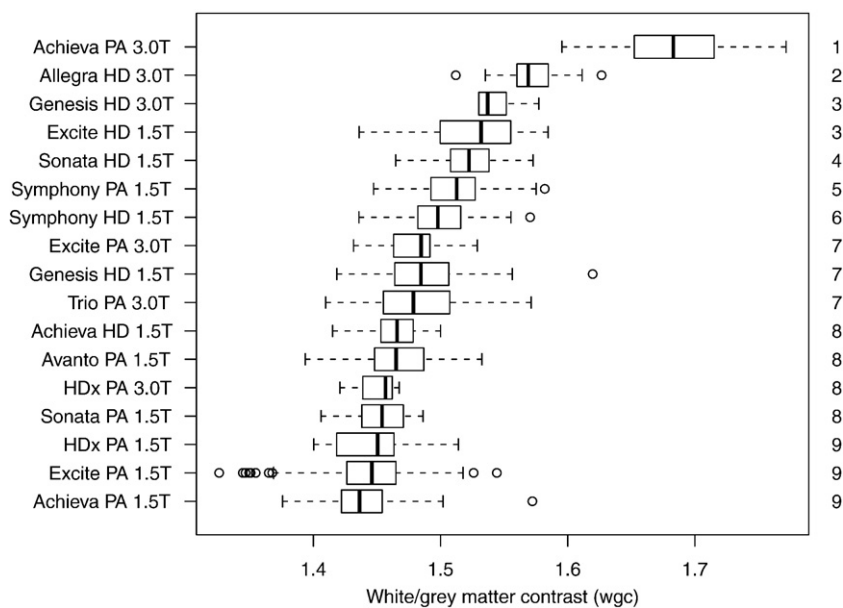
Fig. 4. White/grey matter contrast ratio (WGC) on scanner hardware, sorted by median. Bars in this boxplot denote the median, boxes the 25%/75% quartiles, whiskers the minimum/maximum range. Rankings are shown on the right. Ties denote non-significant differences ($p = 0.05$). Refer to Table 1 for device parameters and a description of the subject sample.

Table 3
Dependency of compartment volumes and ratios on subject and protocol parameters.

|	icv (ml)	brv (ml)	gmv (ml)	wmv (ml)	brr (%)	gwr (%)
Sample mean	1616.0	1195.3	555.4	604.7	71.5	93.9
Age (years)	n.s	−3.48	−1.21	−2.27	−0.170	0.150
Gender (f–m)	142.3	106.0	41.3	65.0	n.s	−3.45
Group (AD-MCI)	n.s	18.7	13.3	n.s	0.856	1.33
Group (AD-Normal)	n.s	38.2	30.8	n.s	2.16	4.29
Body weight (kg)	2.78	1.64	0.802	0.888	−0.015	n.s
Field strength (T)	n.s	n.s	30.5	−24.0	0.820	9.89
WM/GM contrast ratio	n.s	−406.0	−111.4	−324.1	−27.5	32.8
Exp. variance (%)	44.2	48.0	48.1	59.9	55.0	68.6

Data in the first line are grand means in the sample, and may be used to relate to the regression coefficients in the subsequent lines (entries n.s. denote insignificant coefficients). Entries for categorial factors (gender, group) are differences "from-to", e.g., males have an ICV that is 142.3 ml larger.

absolute brain volume *brv* is larger in males, while the relative brain volume *brr* normalizes against body weight and is gender-independent. There is an age-related loss of brain tissue of about 0.35%/year which is stronger in WM than GM. Degenerative processes in WM lead to a signal decrease in T1-weighted images, and the intensity-based segmentation used here may address any lesions to the GM compartment. This explanation is supported by the finding that the GM loss but not the WM loss is dependent on the clinical group. Compared to the normal group, MCI patients have a loss −1.56% in brain volume, and AD patients a loss of −3.20%.

A major device- and protocol-related influence on compartment volumes is the contrast ratio *wgc*. A change in the average contrast of 1.48 by 6% (corresponding to the standard deviation of *wgc* in the sample) leads to a change in computed brain volume by 24 ml (or 2%). Other intensity parameters (e.g., GM or WM intensity) or contrast parameters (e.g., SNR, CNR) may replace *wgc* here, albeit at a lower significance level. Differences between scanner hardware were not significant, except for Excite PA and HDx PA scanners. The GM/WM ratio *gwr* is −19% lower in these systems, corresponding to a lower *gmv* of −7.5%, a higher *wmv* of +11.5%, and a total increase in the brain ratio *brr* by 2.7%.

The impact of scanner type on compartment volumes was studied further by focusing on two systems, Excite PA and Achieva PA, for which 426 examinations were available at 1.5T and 3.0T (see Table 1). The WM/GM contrast ratio on both systems is similar at 1.5T, and almost independent of field strength on the Excite system ($+0.024$/T), in contrast to the Achieva system ($+0.161$/T). This field-dependent effect could not be explained by differences in TR and TE settings alone. The average brain volume is similar at 1.5T (Excite: 1194 ml, Achieva: 1170 ml, $p = 0.07$), and only slightly different at 3.0T (Excite: 1148 ml, Achieva: 1073 ml, $p = 0.02$). The GMV is (roughly) similar on all systems (Excite: 513 ml (1.5T), 547 ml (3.0T), Achieva: 566 ml (1.5T), 577 ml (3.0T), n.s.), but the WMV differs strongly (Excite: 681 ml (1.5T), 601 ml (3.0T), Achieva: 604 ml (1.5T), 496 ml (3.0T), $p = 0$). The difference in WMV largely explains the difference in the GM/WM ratio described above.

Inspection reveals that the higher WM/GM contrast on Philips systems at 3.0T leads to a better delineation of the grey/white matter boundaries. An example is shown in Fig. 5: The same subject is examined on an Achieva PA 3.0T (top) and an Excite PA 1.5T system (below).
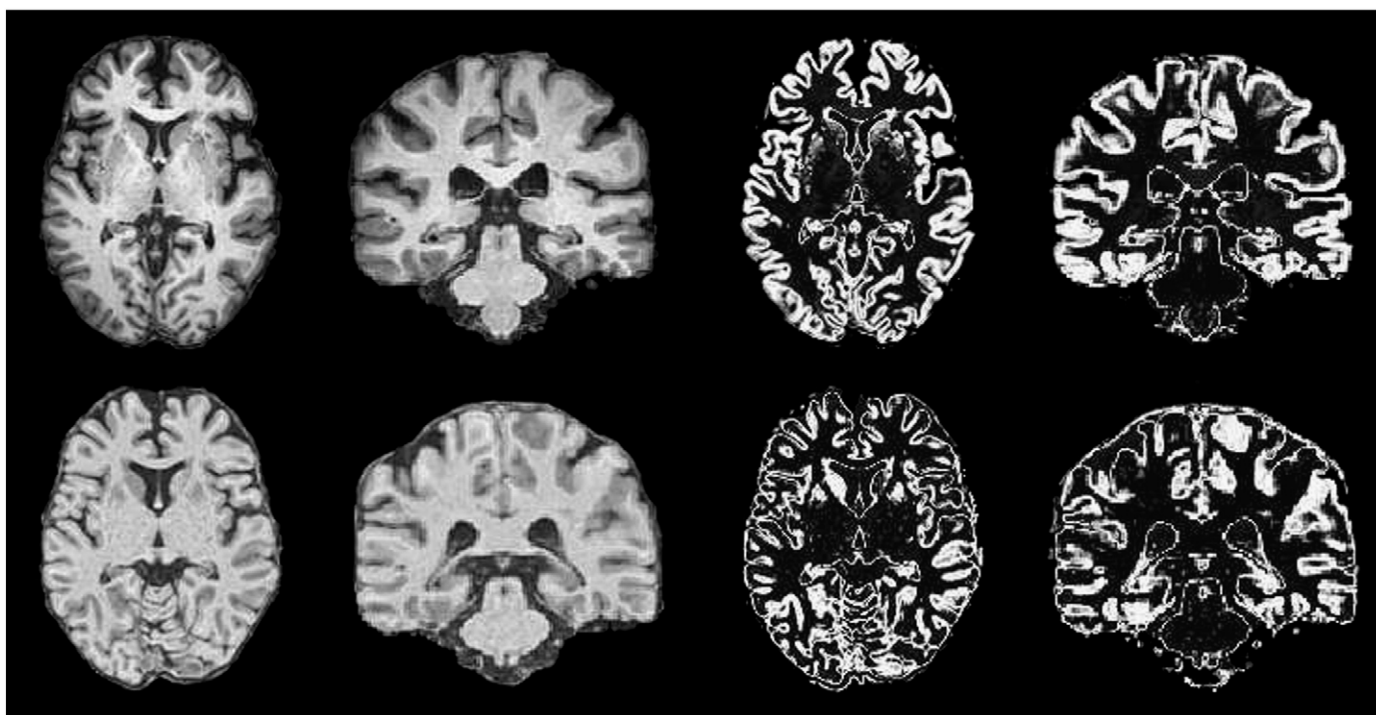
Fig. 5. Axial (column 1) and coronal (column 2) sections of the same subject, examined on a Achieva PA 3.0T (top) and Excite PA 1.5T system (below). Columns 3 and 4 show the corresponding probability images of the GM class. A better WM/GM contrast on the Achieva system leads to a better delineation of the GM cortical layer.

Except for GE scanners, compartment volumes and ratios are similar over all examinations, and thus, independent of the scanning protocol. GE scanners yield larger compartment volumes and a much lower GM/WM ratio. The well-described age-, gender- and group-related influence on compartment volumes and ratios is replicated and confirmed here. Including protocol-related factors when analyzing compartment volumes and ratios yields regression models that explain a much higher proportion of the variance, and leads to more precise estimates of regression coefficients with tighter error bounds. Now, we render these findings more precisely by analyzing results of repeated examinations of the same subject on the same and on different scanners.

Compartment volume intra-scanner variability

A group of subjects were scanned on the same device using the same protocol within a short timeframe (on average 30 days). Of the 43 subjects in the database with repeated scans, two were excluded for quality annotations. Scanner hardware, sites and demographic data of the remaining 82 examinations in 41 subjects are compiled in Table 4, columns "Retest same scanner".

Volumetric data and ratios of paired examinations were converted into within-subject variability by dividing the absolute within-subject difference by the within-subject mean for a given parameter d, expressed in percent: $d_{var} = 200|d_2 - d_1|/d_1 + d_2$, where d_1 corresponds to a measure obtained in examination 1, and d_2 to the result of the second examination. The within-subject variability of the compartment volumes *icv*, *brv*, *gmv*, *wmv* and the ratios *brr*, *gwr* did not depend on scanner hardware, protocol parameters and subject variables, except for the contrast ratio *wgc* that had a weak influence ($p = 0.014$) on the variability of GM and WM volumes. Quantiles of the parameter distributions were determined for all variability measures and are compiled in Table 5. Although absolute differences are not normally distributed, we included the standard deviation (in %, relative to the mean) for informational purposes.

This examination-dependent variability of the volumetric and ratio measures can be used as a lower error bound in longitudinal studies. For example, the standard deviation of the intra-subject difference in the brain ratio is 0.21%. Thus, a change in the brain ratio of 0.42% in a longitudinal study of a single subject may be considered as significant based on an error probability of 5%. Comparing this figure with the overall age-related decrease in *brr* of −0.17%/year, longitudinal changes become significantly detectable after 3 years.

Table 5
Absolute within-subject variability (in %) of compartment volumes and ratios for repeated scans on the same scanner. Quantiles of the distributions are tabulated.

Parameter	Quantile			Std. Dev.
	10%	50%	90%	(%)
Intracranial volume	0.037	0.286	1.27	0.49
Brain volume	0.031	0.296	1.09	0.47
GM volume	0.257	0.788	2.49	0.82
WM volume	0.181	0.088	1.40	0.69
Brain ratio	0.061	0.191	0.57	0.21
GM/WM ratio	0.374	1.288	3.51	1.33
WM/GM contrast	0.064	0.403	1.24	0.76

The higher variability in GMV and WMV is explained by the influence of the contrast on the segmentation: a greater contrast results in a lower variability of GMV and WMV estimates.

Sorting the median of the parameters included in Table 5 by clinical status, a typical ordering of Normal<MCI<AD was found, i.e., normal controls have a better retest reliability. However, differences between groups are not significant (Wilcoxon rank sum test, $p = 0.05$). Likewise, the retest reliability is statistically not significantly different across scanners. Note that the number of subjects per system is small for most scanner types, so this result should be taken with care.

Compartment volume inter-scanner variability

A group of subjects were scanned on different scanners within a short timeframe (on average 30 days). Scanner hardware, sites and demographic data of 344 examinations in 172 subjects are compiled in Table 4, columns 9–15. Paired results were converted into within-subject variability as described in the previous section are compiled as quantiles in Table 6).

Comparing with Table 5, a striking difference is revealed: intra-subject variances are an order of magnitude higher in cross-scan conditions than in repeated scans under the same conditions. If different scanners on the order of those seen here (a 1.5T to 3T upgrade, for example) are used in a longitudinal study, a change in brain ratio of 7.8% is necessary based on a significance level of 5%. This amount corresponds to the expected loss of brain volume in a healthy population over 30 years (Kruggel, 2006). Now, we re-examine the scanner impact of compartment volumes initially described in section "Impact of imaging protocol on compartment volume precision".

Table 4
Retest examinations on the same scanner (columns 2–8) and on different scanners (columns 9–15), detailed per scanner type by number of sites (columns 2, 9), number of examinations (columns 3–5, 10–12), and clinical status (columns 6–8, 13–15).

Scanner hardware	Retest same scanner							Retest different scanner						
	Sites	∑	♀	♂	Normal	MCI	AD	Sites	∑	♀	♂	Normal	MCI	AD
Excite HD 1.5T	2	6	0	6	2	0	4	1	2	0	2	0	1	1
Excite PA 1.5T	5	20	4	16	6	8	6	10	82	42	40	27	36	19
Excite PA 3.0T	0	0	0	0	0	0	0	4	17	10	7	8	6	3
Genesis HD 1.5T	1	4	0	4	0	4	0	3	14	7	7	3	7	4
Genesis HD 3.0T	0	0	0	0	0	0	0	1	6	3	3	2	2	2
HDx PA 1.5T	0	0	0	0	0	0	0	1	3	2	1	2	0	1
HDx PA 3.0T	0	0	0	0	0	0	0	1	3	0	3	2	1	0
Achieva HD 1.5T	0	2	0	2	0	3	0	1	5	3	2	3	1	1
Achieva PA 1.5T	4	10	6	4	4	6	0	4	18	4	14	6	9	3
Achieva PA 3.0T	0	0	0	0	0	0	0	8	44	19	25	15	19	10
Allegra HD 3.0T	0	0	0	0	0	0	0	3	29	16	13	10	14	5
Avanto PA 1.5T	1	4	0	4	2	2	0	2	3	1	2	2	1	0
Sonata HD 1.5T	0	0	0	0	0	0	0	2	8	6	2	2	1	5
Sonata PA 1.5T	1	2	2	0	0	2	0	1	5	3	2	1	4	0
Symphony HD 1.5T	0	0	0	0	0	0	0	2	25	14	11	6	15	4
Symphony PA 1.5T	2	4	4	0	0	2	2	2	7	4	3	2	5	0
Trio PA 3.0T	1	30	16	14	6	18	6	13	73	38	35	17	38	18
Totals	17	82	32	50	18	46	18	59	344	172	172	108	160	76

Note that each subject was scanned twice, i.e., each subject is counted twice in this table. Some subjects took part in both retests, so some of the examinations are counted twice.

Table 6
Absolute within-subject variability (in %) of compartment volumes and ratios for repeated scans on different scanners. Quantiles of the distributions are tabulated.

Parameter	Quantile			Std. Dev.
	10%	50%	90%	(%)
Intracranial volume	0.47	1.66	3.92	1.57
Brain volume	0.65	4.76	12.34	4.64
GM volume	1.77	7.80	20.9	7.11
WM volume	4.80	15.2	32.9	11.9
Brain ratio	0.42	3.64	10.2	3.91
GM/WM ratio	10.43	22.7	50.5	16.4
WM/GM contrast	0.64	3.97	16.9	6.52

To allow a fair comparison, systems with less than 10 examinations were excluded (refer to Table 4). To separate within- and between-subject variability, linear mixed effect models (Baayen et al., 2008) were computed, using age, gender and body weight as covariates, and subject as random factor into account. Results are compiled in Table 7.

Differences in brain compartment volumes of the same subject scanned on different systems are best understood by remembering that compartments have two large boundaries, the WM/GM and GM/CSF interface. A minute shift of 0.1 mm in a cortex of 3 mm thickness leads to a change in the cortical volume of about 3% (or 20 ml). The direction of the differences in compartment volumes found for the Excite PA 1.5T system (see Table 7) can be explained by a relative boundary shift outwards from the WM to the GM and the GM to CSF, leading to a relative increase in wmv and brv at the expense of the GM compartment. The opposite effect is seen on Achieva PA 3.0T and Allegra HD 3.0T systems, with a relative decrease in wmv and brv, without affecting gmv. A shift between the GM/CSF boundary is found on Genesis HD 1.5T, resulting in a decrease in brv and gmv. Finally, a shift of the GM/WM boundary explains results obtained on Trio PA 3.0T systems, with an increase of gmv at the expense of the WM compartment. These boundary shifts also explain the differences found in the brain ratio (d_{brr}) and grey/white matter volume ratio (d_{gwr}).

The GM/WM volume ratio gwr vs. the brain ratio brr is plotted for these scanner types in Fig. 6. Solid ellipses correspond to the within-scanner variance for repeated scans of the same subject on the same scanner; dotted ellipses correspond to the variance pooled across site with the same scanner hardware, corrected for influences of age, gender, clinical status, and body weight. Note that gwr is about 20% lower on Excite PA 1.5T, and about 20% higher on Achieva PA 3.0T, Allegra HD 3.0T, and Trio PA 3.0T. These scanner-related differences explain the large intra-subject variances found for retests of the same subject on different scanners.

A likely reason for these shifts between compartments are the remarkable differences in the tissue contrast. Two 3.0T systems with the highest tissue contrast (Achieva PA, Allegra HD) show similar differences in compartment volumes, while the Excite PA 1.5T system has a low tissue contrast and shows the opposite differences. The much higher variance in compartment volumes across sites than
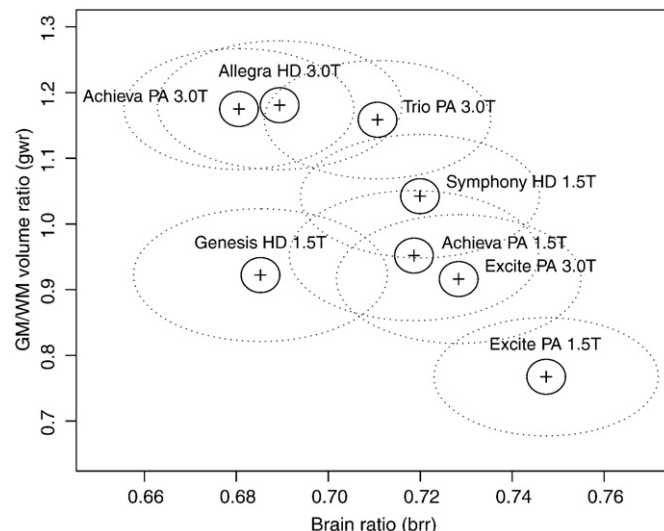
Fig. 6. Grey/white matter volume ratio vs. brain ratio for different scanner hardware. Solid ellipses correspond to the 2σ variance for repeated scans on the same scanner, dotted ellipses correspond to the 2σ variance on the same scanner hardware, pooled across sites. Refer to Table 4 for the number of sites and examinations per scanner type.

within the same system is best explained by differences in the geometrical mapping of scanners.

Sorting the median of the parameters included in Table 5 by clinical status as described in the previous section does not lead to a typical ordering, because scanner-dependent influences on compartment volumes and ratios are much larger than disease-related changes.

Summarizing, the intra-subject variability of compartment volumes and ratios for scans on different systems is roughly 10 times higher than repeated scans on the same system. Possible factors explaining this higher variability are scanner-dependent geometrical inaccuracies and protocol-related differences in tissue contrast, resulting in differences in GM/WM volume ratios.

Discussion

This is the first report of a quantitative assessment of image and segmentation precision in anatomical MR brain imaging based on a large scale multicenter study. We analyzed data acquired by the Alzheimer's Disease Neuroimaging Initiative (ADNI) for their longitudinal study and focused on the baseline examination. Building on the experience of previous multicenter studies in anatomical imaging (e.g., (van Haren et al., 2003; Turner et al., 2006)), the ADNI group took great care to define an optimized mandatory MPRAGE imaging protocol across all sites (Jack et al., 2008).

Although factors influencing MR image quality are well known, this study provides a quantitative analysis of multicenter data. The questions motivating this study are: (1) To what extent do parameters of the imaging protocol (e.g., scanner hardware, TR, TE, voxel size) influence image quality (e.g., SNR, CNR, MI, tissue contrast)? (2) How

Table 7
Within-subject variability of compartment volumes and ratios for repeated scans on different scanners.

Scanner hardware	d_{icv} (ml)	d_{brv} (ml)	d_{gmv} (ml)	d_{wmv} (ml)	d_{brr} (%)	d_{gwr} (%)	d_{wgc} (%)
Excite PA 1.5T	−73.4*	31.2*	−51.8***	87.2***	2.65***	−19.7***	n.s.
Excite PA 3.0T	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Genesis HD 1.5T	n.s.	−54.5***	−44.0**	n.s.	−2.72***	n.s.	n.s.
Achieva PA 3.0T	−46.0**	−117.0***	n.s.	−109.4***	−5.18***	21.2***	23.5***
Allegra HD 3.0T	n.s.	−56.4***	n.s.	−79.2***	−2.43***	21.2***	11.8***
Symphony HD 1.5T	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	5.0**
Trio PA 3.0T	n.s.	n.s.	48.0**	−47.3**	n.s.	19.0***	3.2*

Results correspond to differences in the respective variable, as determined by linear mixed effects models using age, gender, clinical status, and body weight as covariates. Only significant results for scanner hardware with more than 10 examinations are shown ($p < 0.05$: *, $p < 0.01$: **, $p < 0.001$: ***).

much do imaging protocol and quality parameters affect segmentation results (e.g., compartment volumes and ratios)? (3) Is it possible to trace longitudinal changes in single subjects, allowing an individual risk assessment?

We summarize our results as follows:

- We categorize scanner hardware *sh* as the combination of device, coil type and field strength. Due to the optimized imaging protocol in this study, protocol parameters repetition time *tr*, echo time *te* and voxel size *vs* are highly correlated with the scanner hardware *sh*. This parameter typically explains 30–50% of the variance of any independent variable studied here. Including *sh* as covariate improved regression models enormously, and led to tighter error bounds on (other) interesting dependent factors. In conclusion, it is important to include *sh* in statistical models when analyzing multicenter MRI data.

- Despite the standardized protocol, differences across scanners in image quality parameters are considerable, and readily confirmed by visual inspection. Array coils (PA) offer a remarkable advantage over conventional coils (HD) in terms of absolute image noise, and thus, result in a better SNR and CNR. High-field systems generally offer a higher WM/GM contrast ratio, and HD coils have a slight advantage over PA coils here. In summary, systems with PA coils and 3.0T field strength typically rank highest in terms of imaging quality.

- It is well known that absolute volumes of head compartments are correlated with body size, and thus, with gender (Kruggel, 2006). Although body height measures were not available here, body weight was used as a presumably weaker correlate. It is well understood that an imaging study can hardly be balanced for gender and clinical group vs. scanner hardware. As a consequence, body weight was an important confound on all absolute compartment measures. Normalizing against the intracranial volume largely removes the weight- and gender dependence. Because the brain ratio has a maximum of 87.6% during the third decade of life, its actual value may also serve to estimate the overall brain atrophy. Normalizing against the brain volume is not advised, because an atrophy in specific structures (e.g., the temporal lobe) will result in a positive bias in unaffected structures.

The well-described age-, gender- and group-related influences on compartment volumes and ratios are replicated and confirmed here (Ewers et al., 2005; Hua et al., 2008; Kruggel, 2006). Because atrophy rates are small (-0.17%/year), including protocol-related factors is important to find more precise estimates, typically with higher significance. No significant differences in imaging quality across clinical groups were found.

- Repeated scans of the same subject under the same protocol allowed estimating the precision of compartment measures, e.g., for the intracranial volume of 0.49%, for the brain ratio of 0.21% (see Table 5). Relating this figure to the age-related atrophy, longitudinal changes become significantly detectable in individual subjects after 3 years. A higher variability in GMV and WMV is most likely due to an influence of the WM/GM contrast on the segmentation: datasets with a higher contrast had a lower variability. Besides the scanner device, other protocol variables and quality parameters had little influence on compartment measures (Shuter et al., 2008).

- Repeated scans of the same subject on different scanners revealed that precision of compartment measures is roughly 10 times worse than on the same scanner over time (Ewers et al., 2005; van Haren et al., 2003; Schnack et al., 2005). Reconsider Fig. 6, where the GM/WM volume ratio *gwr* is plotted against the brain ratio *brr* for different scanner hardware. The 2σ range of the same-scanner variability is indicated by solid ellipses, and dotted ellipses denote

the within-scanner variability across different sites—the spread across systems is remarkable. Possible explanations for this higher variability are differences in the geometrical mapping of scanners and protocol-related differences in the WM/GM contrast, resulting in much different GM/WM volume ratios. Geometrical errors may be corrected using phantom measurements. The standard ADNI processing pipeline involves a phantom-based geometric correction (Jack et al., 2008) that was not applied in this study to provide an unbiased view. Initial results of this correction (Clarkson et al., 2009) applied to repeated scans of the same patient under the same imaging conditions revealed a volumetric correction of much less than 1%, in line with our findings. However, correcting for differences in gwr due to the imaging protocol is difficult and arguable. We selected a subject who was scanned twice on a Excite PA 1.5T system, and once on a Allegra HD 3.0T system, registered the intensity-corrected images using linear registration (including scaling) to a common reference, and subtracted image pairs (see Fig. 7). While the within-system difference (top) is small, the across-system difference (below) is not only due to a different WM/GM contrast: note the white lines along the WM/GM boundary that indicate a shift on this boundary across systems.

The considerable differences between systems renders pooling absolute measures and ratios currently as arguable. Due to the much lower within-system variability, repeating scans under the same conditions is strongly advised for a longitudinal study. It is expected that relative measures, e.g., atrophy rates, are comparable across systems. This is in line with findings of a multicenter study using legacy data (Fennema-Notestine et al., 2007). Some of the unexplained differences in results of previous morphometric studies may be understood as scanner- and protocol-related.

We are aware of several methodological issues in our study. Most parameters studied here are given (e.g., subject- and protocol-related variables) or result from simple calculation over large samples (e.g., SNR, CNR, MI, WM/GM contrast) and are considered as robust. In
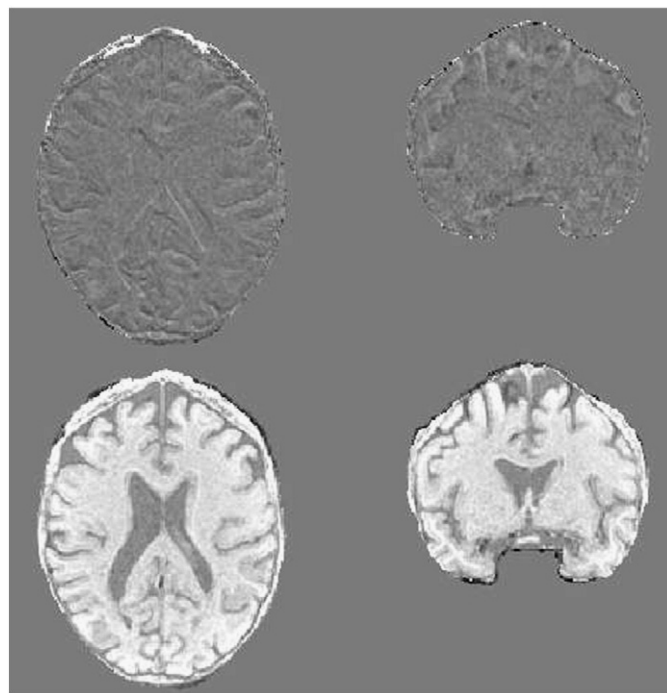
Fig. 7. A subject was scanned twice on a Excite PA 1.5T system, and once on a Allegra HD 3.0T system. Data were linearly registered to a common reference and subtracted to yield the within-system difference (top), and the across-system difference (below). Note the much larger difference due to a different WM/GM contrast and the intense lines along compartment boundaries that indicate a system-related shift.

order to make an unbiased comparison possible, we used a fairly simple and automated image segmentation chain. All images were processed without possible optimization towards a specific imaging protocol, and it may be possible to reduce some of the variability if a scanner-optimized processing is used. The low variability of 0.2–0.7% in the volumetric measures from repeated scans using the same protocol is well in line with results of other studies (Han et al., 2006; Kruggel, 2006). The much higher variance across protocols may be considered as a lack in robustness of the segmentation procedure. However, a major portion of this variance was explained in terms of protocol-dependent parameters (e.g., device and field strength) or robust measures such as the WM/GM contrast ratio. In addition, image quality measures (SNR, CNR, MI) proved to have little influence on compartment volume estimates, indicating a considerable stability of the segmentation algorithm. For an intensity-based method as employed here, the difficulty of a segmentation problem can be recast in terms of a tissue separation measure: the white-to-grey matter intensity difference divided by the sum of the intraclass variances. The segmentation is easier in images as shown in Fig. 5 (top), and rather difficult in images as shown below.

Noise levels are best estimated in regions of homogeneous media. In MR images of the human head, unfortunately, no tissue compartment can be considered as homogeneous: (1) the white matter and the ventricles show texture that is related to structure at or below the resolution level of MR imaging; (2) a typical finding in elderly subjects are diffuse WM lesions; (3) in the background, image imperfections (e.g., ghost images, pulsation artifacts) lead to local signal changes. We have chosen to estimate noise in the whole domain of "raw" images, and consider that our noise measure rather over-estimates the true noise level.

Coil properties are not uniform in space. As a consequence, the level of absolute noise may not be uniform in space. Systems equipped with phased-array coils apply scanner-internal correction mechanism that are largely opaque to the end user. Since we do not know exactly how the scanner software reconstructed (combined) the images from the phased array elements, it is hard to predict if our integral noise measure is an under- or over-estimation of the true noise level (refer to Roemer et al., 1990 for an in-depth discussion). Thus, we cannot estimate how much scanner-internal corrections of coil properties lead to the advantage of using PA coils as demonstrated here.

To keep this analysis straightforward, we did not compare other segmentation schemes in this paper. We also have processed the data using the segmentation chain for voxel-based morphometry as included in SPM5 (Ashburner and Friston, 2000), but found unplausible results in about 10% of the segmentation results. It is worthwhile and necessary to study the advantage of protocol-specific corrections (e.g., a phantom-based geometrical correction, parameter optimization) (Han and Fischl, 2006; Jovicich et al., 2006). However, the clinically more realistic scenario is that a subject is repeatedly scanned on the same scanner by the same protocol. Because the intra-protocol precision is much higher, and relative measures (e.g., atrophy rates) are of predominant interest, pooling results from multicenter studies still appears viable. However, care is advised when comparing absolute measures.

To avoid a further confound with methodological issues, more disease-relevant measures (e.g., the neocortical thickness, hippocampal volume, (see Han et al., 2006; Leow et al., 2009; Jovicich et al., 2009)) were not studied here, because we expect that influences of the segmentation method play a larger role here. However, to estimate subject sample sizes in a power analysis (Fox et al., 2000), such measures may be more sensitive than the overall brain atrophy rate.

Morphometric methods such as tensor-based morphometry use nonlinear registration to compute a deformation field that describes inter-subject differences. Scanner-dependent differences in geometrical mapping are rather large-scale and will increase the overall variance of the mapping. Thus, this error can and should be corrected (e.g., based on phantom scans). Protocol-dependent shifts in the

definition of compartment boundaries and differences in the WM/GM contrast are harder to correct, and lead to differences in the estimation of compartment volumes, especially of smaller structures such as the basal ganglia (e.g., the white line on the border of the caudate nucleus in Fig. 7, below). If possible, the scanner type (the device-coil-field strength combination used here) should be included as a covariate.

Another possible confound on the GM/WM ratio was not studied systematically here. Besides the well-described age- and disease-related cortical atrophy, degenerative effects acting on the white matter lead to an increase in extracellular volume, and thus, to a diffuse decrease in signal intensity, and to a higher prevalence of diffuse white matter lesions. Such diffuse lesions may be incorrectly classified as "grey matter" here. This effect may explain the higher atrophy rate in WM than in GM found here (see Table 3). More elaborate segmentation schemes (Kruggel et al., 2007) estimate the lesion load that can be used to compute more correct atrophy rates.

It is highly desirable that imaging results are similar across scanner systems. However, a number of issues (e.g., technological advance, patenting) may render this request as unrealistic or even undesirable. Technological advances, e.g., provided by high-field systems equipped with array coils, result in a much better image quality, making the segmentation task easier, and thus, reducing segmentation errors. Results presented here may stimulate the discussion about a better standardization in medical imaging.

Acknowledgments

References

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—the methods. NeuroImage 11, 805–821.

Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59, 390–412.

Bishop, C.M., 1995. Neural networks for pattern recognition. Claredon Press, Oxford, p. 1995.

Clarkson, M.J., Ourselin, S., Nielsen, C., Leung, K.K., Barnes, J., Whitwell, J.L., Gunter, J.L., Hill, D.L., Weiner, M.W., Jack, C.R., Fox, N.C., 2009. The Alzheimer's disease neuroimaging Initiative. Comparison of phantom and registration scaling corrections using the ADNI cohort. NeuroImage 47, 1506–1513.

Ewers, M., Teipel, S.J., Dietrich, O., Sch'onberg, S.O., Jessen, F., Heun, R., Scheltens, P., van de Pold, L., Freymann, N.R., Moeller, H.J., Hampel, H., 2005. Multicenter assessment of reliability of cranial MRI. Neurobiol. Aging 27, 1051–1059.

Fennema-Notestine, C., Gamst, A.C., Quinn, B.T., Pacheco, J., Jernigan, T.L., Thal, L., Bucker, R., Killiany, R., Blacker, D., Dale, A.M., Fischl, B., Dickerson, B., Gollub, R.L., 2007. Feasibility of multi-site clinical structural neuroimaging studies of aging using legacy data. Neuroinform 5, 235–245.

Fox, N.C., Cousens, S., Scahill, R., Harvey, R.J., Rossor, M.N., 2000. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer's disease: power calculations and estimates of sample size to detect treatment effects. Arch. Neurol. 57, 339–344.

Gudbjartsson, H., Patz, S., 1995. The Rician distribution of noisy MRI data. Magn. Reson. Med. 34, 910–914.

Han, X., Fischl, B., 2006. Atlas renormalization for improved brain MR image segmentation across scanner platforms. IEEE Trans. Med. Imaging 26, 479–486.

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B., Fischl, B., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. NeuroImage 32, 180–194.

Hentschel, S., Kruggel, F., 2004. Segmentation of the intracranial compartment: A registration approach. In: Jiang, T. (Ed.), Medical Imaging and Augmented Reality (Beijing), Lecture Notes in Computer Science, Vol. 3150. Springer, Singapur, pp. 253–260.

Hua, X., Leow, A.D., Lee, S., Klunder, A.D., Toga, A.W., Lepore, N., Chou, Y.Y., Brun, C., Chiang, M.C., Barysheva, M., Jack, C.J., Bernstein, M.A., Britson, P.J., Ward, C.P., Whitwell, J.L., Borowski, B., Fleisher, A.S., Fox, N.C., Boyes, R.G., Barnes, J., Harvey, D., Kornak, J., Schuff, N., Boreta, L., Alexander, G.E., Weiner, M.W., Thompson, P.M., 2008a. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. NeuroImage 41, 19–34.

Hua, X., Leow, A.D., Parikshak, N., Lee, S., Chiang, M.C., Toga, A.W., Jack, C.R., Weiner, M.W., Thompson, P.M., 2008b. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: an MRI study of 676 AD, MCI, and normal subjects. NeuroImage 43, 458–469.

Jack, C.R., Shiung, M.M., Weigand, S.D., O'Brien, P.C., Gunter, J.L., Boeve, B.F., Knopman, D.S., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Petersen, R.C., 2005. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI. Neurology 65, 1227–1231.

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L. G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, S.C., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Med. 27, 685–691.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., Macfall, J., Fischl, B., Dale, A., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. NeuroImage 30, 436–443.

Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B.C., Fischl, B., 2009. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. NeuroImage 46, 177–192.

Kruggel, F., 2006. MRI-based volumetry of head compartments: normative values of healthy adults. NeuroImage 30, 1–11.

Kruggel, F., Lohmann, G., 1996. BRIAN (brain image analysis)—a tool for the analysis of multimodal brain data sets. Proc. Computer Aided Radiology 1996. Springer, Berlin, pp. 323–328.

Kruggel, F., von Cramon, D.Y., 1999. Alignment of magnetic-resonance brain datasets with the stereotactical coordinate system. Med. Imag. Anal. 3, 1–11.

Kruggel, F., Paul, J.S., Gertz, H.J., 2007. Texture-based segmentation of diffuse lesions of the brain's white matter. NeuroImage 39, 987–996.

Leow, A.D., Yanovsky, I., Parikshak, N., Hua, X., Lee, S., Toga, A.W., Jack, C.R., Bernstein, M.A., Britson, P.J., Gunter, J.L., Ward, C.P., Borowski, B., Shaw, L.M., Trojanowski, J.Q., Fleisher, A.S., Harvey, D., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.M., Alzheimer's Disease Neuroimaging Initiative, 2009. Alzheimer's disease neuroimaging initiative: a one-year follow up study using tensor-based morphometry correlating degenerative rates, biomarkers and cognition. NeuroImage 45, 645–655.

Mortamet, B., Bernstein, M.A., Jack, C.R., Gunter, J.L., Ward, C., Britson, P.J., Meuli, R., Thiran, J.P., Krueger, G., Alzheimer's Disease Neuroimaging Initiative, 2009. Automatic quality assessment in structural brain magnetic resonance imaging. Magn. Reson. Med. 62, 365–372.

Mueller, S.G., Weiner, M.W., Thal, L.J., et al., 2005. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin. North Am. 15, 869–877.

Pham, D.L. 2001 Robust fuzzy segmentation of magnetic resonance images. In: Proc. 14th IEEE Symposium on Computer-Based Medical Systems (CBMS2001), pp. 127-131.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. Numerical Recipes. 3rd Edition. Cambridge University Press, Cambridge.

Roemer, P.B., Edelstein, W.A., Hayes, C.E., Souza, S.P., Mueller, O.M., 1990. The NMR phased array. Magn. Reson. Med. 16, 192–225.

Schnack, H.G., van Haren, N.E.M., Hulshoff Pol, H.E., Picchioni, M., Weisbrod, M., Sauer, H., Cannon, T., Huttunen, M., Murray, R., Kahn, R.S., 2005. Reliability of brain volumes from multicenter MRI acquisition: a calibration study. Hum. Brain Mapp. 22, 312–320.

Shuter, B., Yeh, I.B., Graham, S., Aub, C., Wang, S.C., 2008. Reproducibility of brain tissue volumes in longitudinal studies: effects of changes in signal-to-noise ratio and scanner software. NeuroImage 41, 371–379.

Turner, J.A., Smyth, P., Macciardi, F., Fallon, J.H., Kennedy, J.L., Potkin, S.G., 2006. Imaging phenotypes and genotypes in schizophrenia. Neuroinformatics 4, 21–49.

van Haren, N.E.M., Cahn, W., Hulshoff Pol, H.E., Schnack, H.G., Caspers, E., Lemstra, A., Sitskoorn, M.M., Wiersma, D., van den Bosch, J.R., Dingemans, P.M., Schene, A.H., Kahn, R.S., 2003. Brain volumes as predictor of outcome in recent-onset schizophrenia: a multi-center MRI study. Schizophr. Res. 64, 41–52.