



Published in final edited form as:

Neuroimage. 2012 April 2; 60(2): 940–951. doi:10.1016/j.neuroimage.2012.01.092.

Combining atlas-based parcellation of regional brain data acquired across scanners at 1.5T and 3.0T field strengths

Adolf Pfefferbaum^{1,2}, Torsten Rohlfing¹, Margaret J. Rosenbloom^{1,2}, and Edith V. Sullivan²

¹Neuroscience Program, SRI International, Menlo Park, CA

²Department of Psychiatry & Behavioral Sciences, Stanford University School of Medicine, Stanford, CA

Abstract

Longitudinal brain morphometric studies designed for data acquisition at a single MRI field strength can be seriously limited by system replacements from lower to higher field strength. Merging data across field strengths has not been endorsed for a variety of reasons, yet the ability to combine such data would broaden longitudinal investigations. To determine whether structural T1-weighted MRI data acquired across MR field strengths could be merged, parcellations of archival SPGR data acquired in 114 individuals at 1.5T and at 3.0T within 3 weeks of each other were compared. The first set of analyses examined 1) the correspondence between regional tissue volumes derived from data collected at 1.5T and 3.0T and 2) whether there were systematic differences for which a correction factor could be determined and applied to improve measurement agreement. Comparability of regional volume determination at 1.5T and 3.0T was assessed with intraclass correlation (ICC) computed on volumes derived from the automated and unsupervised SRI24 atlas registration and parcellation method. A second set of analyses measured the reliability of the registration and quantification using the same approach on longitudinal data acquired in 69 healthy adults at a single field strength, 1.5T, at an interval <2 years. The mainstay of the analyses was based on the SRI24 method; to examine the potential of merging data across field strengths and across image analysis packages, a secondary set of analyses used FreeSurfer instead of the SRI24 method. For both methods, a regression-based linear correction function significantly improved correspondence. The results indicated high correspondence between most selected cortical, subcortical, and CSF-filled spaces; correspondence was lowest in the globus pallidus, a region rich in iron, which in turn has a considerable field-dependent effect on signal intensity. Thus, the application of a regression-based correction function that improved the correspondence in regional volume estimations argues well for the proposition that selected T1-weighted regional anatomical brain data can be reliably combined across 1.5T and 3.0T field strengths with the application of an appropriate correction procedure.

© 2012 Elsevier Inc. All rights reserved.

Correspondence: Adolf Pfefferbaum, M.D., Director, Neuroscience Program, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, 650-859-2927 (office), 650-859-2743 (FAX), dolf@synapse.sri.com.

Statement on conflict of interest

No author of this manuscript has any conflicts of interest with this work, either financial or otherwise.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

longitudinal; MRI; brain; field strength; atlas-based parcellation; image registration; SRI24; FreeSurfer

Introduction

Over the last two decades, quantitative neuroimaging studies using magnetic resonance imaging (MRI) have been conducted principally at 1.5T field strength. The availability of higher field strength, notably 3.0T, for clinical research purposes at times has replaced 1.5T systems, thereby derailing studies designed to be conducted at a single field strength. Even in centers that have maintained the lower field strength systems along with higher-strength systems, the availability of both poses a dilemma for longitudinal neuroimaging studies, especially those initiated on 1.5T systems: should a study be continued at the lower field strength to ensure longitudinal comparability of signal and quantification; alternatively, should a study change mid-stream to the higher field strength system with higher signal-to-noise and contrast-to-noise ratio per imaging time? Merging data across field strengths, however, has not been endorsed for a variety of reasons, including differences in field strength effects from B1-field homogeneity on various brain tissue types and their locations in the magnetic field. Yet, the ability to combine 1.5T and 3.0T data could expand sample sizes and holds other possibilities for broadening longitudinal investigations of changes in regional brain morphology and tissue integrity associated with normal development, normal aging, disease progression, and spontaneous or therapeutic resolution of disease.

Some MRI studies conducted across field strengths have taken advantage of the differences in relaxivity properties of tissue, especially those related to detection of iron (Duyn, 2011), which occurs with stroke (Srinivasan et al., 2006), some degenerative disorders (Bartzokis et al., 1999; Bartzokis et al., 2007a; Bartzokis et al., 2000), and diffuse axonal injury (Luccichenti et al., 2010), and accrues in selective basal ganglia and brain stem structures with age (Bartzokis et al., 2007b; Pfefferbaum et al., 2009; Pfefferbaum et al., 2010). Signal alteration of iron-laden tissue becomes greater with increasing field strength due to the ferromagnetic property of the iron causing increased T1, T2, and T2* relaxivity. Other comparisons note greater enhancement of a target using contrast agents, for example, to detect brain tumors; an early study showed greater enhancement at higher field strength, in this case 2.0T relative to 0.5T (Chang et al., 1994), and other more recent work corroborates this general finding (Attenberger et al.; Pinker et al., 2008). Other studies indicate conflicting findings related to tissue composition and pathological markers. In comparing the utility of 1.5T compared with 3.0T on clinical reading in detecting lesions in presurgical epilepsy patients, the two reviewing neuroradiologists had higher inter-rater agreement on the 3.0T than 1.5T images but identified more lesions at 1.5T than at 3.0T, and the lesions identified were different between scanners (Zijlmans et al., 2009). Differences noted in these and other studies known to influence tissue signal and border conspicuity (e.g., Bammer et al., 2007; Boss et al., 2007; Fushimi et al., 2007; Stankiewicz et al., 2011; Zhu et al., 2011) militate against combining data across field strengths for equivalent measurement of structural brain volumes unless the differences were detectable, measurable, regular, and therefore amenable to correction.

Several attempts have recently been made to assess the possibility of combining data across 1.5T and 3.0T field strengths. One study focused on measurement of intracranial volume (ICV) because of its utility in normalizing variability in regional volume estimates attributable to ICV differences (Keihaninejad et al., 2010). After correcting for B1-field inhomogeneity and applying a reverse normalization strategy for registration, notable at

3.0T, measurement reliability between magnet strengths was high, despite differences in MR system manufacturers and head coils. Another study focused on comparability of volume measurements of subcortical brain structure, also across 1.5T and 3.0T systems made by different manufacturers and using different head coils. (Goodro et al., 2011). Because different subjects were scanned on the different MR systems, the comparison was based on an outcome measure, which was the similarity of relations between structural volumes and age. The high agreement between volume estimates suggested comparability of the scanners in identifying effects of normal aging on subcortical measures. Cortical thickness examined with 1.5T and 3.0T scanners resulted in highly similar measurements with 3.0T providing slightly higher thickness estimates in the same 15 subjects scanned multiple times (Han et al., 2006). A further analysis of these data focused on regional brain volumes and indicated higher reliability across field strengths within the same manufacturer's platform than between different manufacturer's platforms (Jovicich et al., 2009).

In an effort to determine whether structural MRI data acquired at 1.5T and 3.0T, the two most widely available magnetic strengths for clinical and research purposes, could be merged, we assembled archival data acquired at 1.5T and at 3.0T in close temporal proximity in the same individuals to address measurement comparability. Accordingly, the first set of analyses examined 1) the correspondence between measurements derived from data collected at 1.5T and 3.0T and 2) whether systematic differences in volume estimates existed for which a linear correction could be determined and applied to improve measurement agreement. Comparability of regional volume determination at 1.5T and 3.0T was assessed using the automated and unsupervised SRI24 atlas registration and parcellation method, a 3-dimensional structural image quantification approach developed in our laboratory (Rohlfing et al., 2010). To ensure that differences between datasets were not a function of a lack of reliability of the parcellation approach, a second set of analyses measured the reliability of the quantification using the same approach on longitudinal data acquired in healthy adults at a single, field strength, 1.5T, at an interval of less than 2 years. The mainstay of the analyses and presentation herein is based on the SRI24 method. To examine the potential of merging data across field strengths and across image analysis packages, a secondary set of analyses used FreeSurfer, a free, widely-used analysis package based on an "inflation" approach and applied here for automated and unsupervised image parcellation and quantification (Fischl et al., 2004b).

Methods

Subjects

The data reported herein were drawn from ongoing longitudinal studies of brain structure in alcoholism, HIV infection, and normal aging. Descriptions of recruitment, screening, and general demographics appear in previous publications (e.g., Pfefferbaum et al., 2007; Pfefferbaum et al., 2006; Rosenbloom et al., 2007). All subjects had provided written informed consent to participate in these studies, which were approved by the Institutional Review Boards of Stanford University and SRI International. Subjects were included in the analysis based solely on the availability of two scans, one at 1.5T and one at 3.0T, separated by no more than 3 weeks, but independent of age, sex, or diagnosis. The inclusion of such a heterogeneous group provided ample anatomical variability to assess the effects of MR field strength differences, the analysis of which was independent of subject characteristics.

1.5T vs. 3.0T comparison group—The subjects with data at the two field strengths were 84 men and 30 women, age 47.5 ± 10.2 years (21 to 66 years). The total group was composed of 31 men and 38 women who were deemed healthy controls, 19 men and 6 women with alcohol dependence, 25 men and 8 women with HIV-infection, and 24 men and

6 women with HIV infection and alcoholism. The average (\pm S.D.) MRI interval for the entire sample was 4.4 ± 5.9 days, range=0 to 21 days. The MRI interval was 2.48 days (0 to 11 days) for the alcoholics and 3.9 days (0 to 21 days) for the HIV/alcoholic group. Recent alcohol consumption was monitored on breathalyzer each scan day, and date of last drink was recorded as part of the lifetime alcohol history and its update. Of the 61 patients with a history of alcoholism (with or without HIV infection), 19 were current drinkers (within the last month) at the time of the first scan, 21 were in early remission (sober for upwards of 1 year), and 22 were in full remission (sober for more than 1 year). No alcohol-dependent participant reported excessive drinking between the scans.

1.5T reliability group—This group comprised 31 men and 38 women, all serving as healthy controls in ongoing studies, who were each scanned twice on the 1.5T system within a 2-year interval. Their average age was 50.2 ± 15.5 years (range=20 to 81 years), and they underwent MRI scanning on average 13.6 ± 4.42 months apart (range=5.8 to 23.2 months).

MRI acquisition

1.5T acquisition—This set of MRI data was collected on a GE 1.5T Signa Twin whole-body system with a quadrature head coil (General Electric Healthcare, Waukesha, WI). Two coronal structural sequences were used for the analysis: a SPOiled Gradient Recalled Echo (SPGR) sequence (TR=25 ms, TE=5 ms, flip angle=30°, matrix=256×192, thick=2 mm, skip=0 mm, 94 slices) and a dual-echo fast spin echo (FSE) sequence (TR=7500 ms, TE1/2=13.5/108.3 ms, matrix = 256×192, thick=4 mm, skip=0 mm, 47 slices).

3.0T acquisition—This set of MR data was collected on a GE 3.0T Signa whole-body system with an 8-channel phased-array head coil. Data were derived from T1-weighted Inversion-Recovery Prepared SPGR images (TR=7 ms, TE=2.2 ms, TI=300 ms, thick=1.25 mm, skip=0 mm, 124 slices) and dual-echo FSE images (TR=8583 ms, TE1/2=13.5/108.3 ms, thick=2.5 mm, skip=0 mm, 62 slices).

SRI24 atlas-based parcellation

For both 1.5T and 3.0T data, all acquired structural images were first corrected for intensity inhomogeneity by applying a second-order polynomial multiplicative bias field computed via entropy minimization (Likar et al., 2001). The late-echo FSE image was corrected using the bias field computed from the corresponding early-echo image to maintain the ratio of early- and late-echo values at each pixel, which keeps quantities derived from this ratio (e.g., T2) invariant. For each subject and each session, the bias-corrected early-echo FSE image was then registered to the biascorrected SPGR image using intensity-based nonrigid image registration (Rohlfing and Maurer, 2003) (<http://nitrc.org/projects/cmtk>). The SPGR, early-echo FSE, and late-echo FSE images were each skull stripped using FSL's Brain Extraction Tool, BET (Smith, 2002). The early- and late-echo brain masks were reformatted into SPGR image space and combined with the SPGR-derived brain mask via label voting (Rohlfing and Maurer, 2005) to form the final SPGR brain mask.

For each subject and each scan session, the skull-stripped SPGR images were registered to the SPGR channel of the SRI24 atlas (Rohlfing et al., 2010) (<http://nitrc.org/projects/sri24>) via nonrigid image registration (Rohlfing and Maurer, 2003). We chose the SRI24 atlas over other available brain templates (e.g., MNI152) because of its ability to discern detailed anatomical structures, which can thus be unambiguously outlined directly in the atlas images without the need to access the images that were used to create the atlas itself. Cortical and subcortical parcellation maps for all subjects and sessions were obtained by reformatting labels maps defined in SRI24 space directly into SPGR image spaces using the subject-to-atlas coordinate transformations.

All bias-corrected and skull-stripped SPGR images were segmented into three tissue compartments (gray matter, white matter, CSF) using FSL's FAST tool (Zhang et al., 2001). As tissue priors to both initialize and guide the classification, we used the tissue probability maps provided with the SRI24 atlas, reformatted into subject SPGR space via the same transformations described above for the atlas-based parcellation.

FreeSurfer registration and atlas-based parcellation

The unprocessed SPGR data for each subject and each scan session were entered into the FreeSurfer image analysis suite for unsupervised cortical surface reconstruction and volumetric segmentation (<http://surfer.nmr.mgh.harvard.edu/>). The FreeSurfer suite did not make use of the FSE data used for skull stripping in the SRI24 analysis. The technical details of these procedures are described in prior publications by others (Dale et al., 1999; Dale and Sereno, 1993; Fischl and Dale, 2000; Fischl et al., 2001; Fischl et al., 2002; Fischl et al., 2004a; Fischl et al., 1999; Fischl et al., 2004b; Han et al., 2006; Jovicich et al., 2006; Morey et al., 2010; Segonne et al., 2004; Wonderlick et al., 2009). FreeSurfer morphometric procedures have been demonstrated to show good test-retest reliability across scanner manufacturers and across field strengths (Han et al., 2006).

Regions of interest (ROIs)

The SRI24 analysis used 24 regions of interest (Figure 1), including cortical, subcortical, and white matter structures as well as cerebrospinal fluid (CSF)-filled spaces, and as was described in our prior work (Pfefferbaum et al., 2011; Pfefferbaum et al., 2006). For each subject and session, gray matter volume was computed for each cortical region, and tissue volume (gray plus white matter) for each subcortical region. Also measured were CSF-filled volumes of the lateral ventricles, third ventricle, and sylvian fissures as well as total ICV and supratentorial volume.

We selected 21 regions from the FreeSurfer atlas with similarly named, although not anatomically identical, brain structures to match the SRI24-parcellated regions (Destrieux et al., 2010).

Statistical analysis

Comparisons of regional volumes at the two MR field strengths and also between two time points at a single field strength were quantified with Pearson correlations (r) and intraclass correlations (ICC) with 95% confidence intervals. Comparisons of change in ICC before vs. after correction were conducted with t-tests. Statistical analyses were conducted with R (<http://www.r-project.org>) and Statview.

Results

Comparison of regional brain volumes at 1.5T and 3.0T

Comparability of regional volume measurements using the SRI24 atlas—The scatter of the regional volumes from the identity line (Figure 2) and the ratio of mean 1.5T/3.0T volume (Table 1) indicated that 13 regions were larger at 3.0T than 1.5T by .3 to 15.0% and 11 were smaller at 3.0T than 1.5T by .2 to 19.1%. Bivariate correlations and ICCs (Table 1) were calculated for each of the 24 regional volumes from the 114 subjects scanned at both field strengths to assess the direction of discrepancies between measurement of each region and the comparability of measurements collected at 1.5T and 3.0T. All correlations were highly significant and ranged from $r=.652$ to $.998$ (mean $r=.895$, median $r=.887$). The ICCs, a preferred test of reliability over simple correlations because of its sensitivity to intercept differences (Shrout, 1998), ranged from $.216$ for the globus pallidus

to .995 for the lateral ventricles (mean ICC=.801; median ICC=.834). According to the categorization of Landis and Koch (1977), 16 were “substantial,” 5 were “moderate,” 2 were “fair,” 1 (globus pallidus) was only “slight.”

To test whether the 1.5T-3.0T differences were due to simple (global) volumetric scaling, we examined the effect of expressing the regional volumes as a percentage of supratentorial volume or as standardized scores regressed on supratentorial volume. This approach did not improve overall ICCs between 1.5T- and 3.0T-derived volumes, and for most regions made them worse.

SRI24 regression correction function (RCF)—To compensate for the measurement discrepancies, we derived a linear regression correction function (RCF) for each region. For each of the 25 regional volumes, the linear fit of 3.0T on 1.5T volume was computed and the slope and intercept were used to transform the 3.0T volume.

$$\text{Volume}_{3.0T_{\text{RCF}}} = (\text{Volume}_{3.0T} - \text{intercept}) / \text{slope}$$

Application of the RCFs significantly improved the ICCs (Wilcoxon $Z=3.615$, $p=.0003$), representing an average ICC improvement of .087. The post-RCF ICCs ranged from .599 to .998 (mean=.888; median=.881). ICCs improved for 17 regions, declined insubstantially for 4 regions, and remained unchanged for 3 regions (Table 1).

To test the robustness and potential utility of the RCFs, we divided the group of 114 subjects into two random samples (*a* and *b*) of 57 subjects each. The two samples were similar in age ($a=48.7\pm 9.2$ years; $b=46.5\pm 11.2$ years; $t(112)=1.01$, n.s.) and distributions of sex ($\chi^2=.41$, n.s.) and diagnosis ($\chi^2=3.90$, n.s.). We calculated regional RCFs for each sample, and then applied the RCFs to the other sample, that is, the RCFs of sample *a* were applied to sample *b* and vice versa. For the *a*-*b* comparison, the ICC of 19 regions improved and 5 declined, which was an average increase of .079 (Wilcoxon $Z=1.947$, $p=.0515$). 21 were in the substantial range (i.e., greater than .80); 2 (putamen and postcentral cortex) were in the moderate range; and the globus pallidus ICC was fair (.585). The *b*-*a* comparisons were similar, whereby 20 improved and 4 got worse ($t(23)=3.180$, $p=.0042$), mean ICC diff=.072; only the globus pallidus (ICC=.454) failed to reach the substantial ICC range (Table 2).

To explore whether the regression correction factors were disproportionately larger in the patient groups than controls, we used one-way ANOVAs to test for group effects in the signed and absolute RCFs differences between the 1.5T and 3.0T corrected volumes. Family-wise Bonferroni correction for 25 ROIs required $p\leq .002$. For the signed RCFs, group differences emerged for the postcentral cortex ($p=.0007$), where the correction was greater for the control than either the HIV or alcoholic groups, and for the posterior cingulum ($p=.0078$), where the correction was greater for the controls than the comorbid group. While statistically significant, the differences were trivial ($<1\text{mm}^3$). In no case did the group differ significantly in absolute differences of RCFs.

In an effort to determine whether a general, region-dependent correction factor could provide as robust an adjustment as our RCF, we used two global correction factor approaches. One computed a simple scalar factor based on the ratio of the mean of supratentorial volume of 1.5T divided by the same for 3.0T. The other computed the regression of supratentorial volume of 3.0T on 1.5T and applied the slope and intercept to each ROI of the 3.0T data, with the intercept for each ROI defined as the proportion of the supratentorial volume intercept/mean 1.5T supratentorial volume multiplied by the mean volume of a given ROI. The global ratio-based factor resulted in improved ICCs relative to

uncorrected volumes for 7 of 24 ROIs: thalamus, globus pallidus, hippocampus, amygdala, insula, anterior-middle cingulum, and third ventricle. The global regression-based factor resulted in improved ICCs relative to uncorrected volumes for 9 of 24 ROIs: the 7 improved by the ratio-based method plus the postcentral cortex and lateral ventricles. Overall, however, the global regression-based factor diminished the ICCs on average by 6.6% ($t(23)=2.391$, $p=.0254$); the simple scale factor diminished the ICCs on average by 4.5% ($t(23)=1.850$, $p=.0772$). By contrast, the individual ROI regression correction procedure improved the ICCs in 17 of the 24 ROIs, an average of 8.7% ($t(23)=3.770$, $p=.001$); the ICC differences in the remaining 6 ROI were 0 to .03%. These results support the conclusion that the proposed individual ROI RCF approach provided a better 1.5T-3.0T volume adjustment than did the global ICC-based volume adjustment approaches.

Comparability of regional volume measurements using FreeSurfer-derived volumes—We selected 21 regions from the FreeSurfer atlas that represented similar brain structures to the SRI24-parcellated regions. The average and range of 1.5T to 3.0T measurement ICCs and correlations were as follows: median ICC=.763, mean ICC=.722, range=.258 to .997; median $r=.877$, mean $r=.860$, range=.542 to .998 (Table 3 and Figure 4).

Application of the RCFs significantly improved all 21 ICCs (Wilcoxon $Z=4.015$, $p=.0001$), representing an average ICC improvement of .126. The post-RCF ICCs ranged from .457 to .998 (mean=.847; median=.870).

Reliability of regional brain volumes over time at 1.5T

The second set of analyses measured the reliability of the registration and quantification using the SRI24 parcellation on longitudinal data acquired in 69 healthy control men and women scanned at a single field strength, 1.5T, at an interval of less than 2 years. To minimize systematic age-related changes over the scan interval, we randomly alternated the temporal order of the two acquisitions so that for half the subjects the earlier scan was chosen as the first observation, and for the other half the second scan was the first observation. The correlations between MRI sessions were highly significant for each regional volume and ranged from $r=.829$ to .996 (mean $r=.949$; median $r=.956$). The ICCs ranged from .830 in the putamen to .996 in the lateral ventricles (mean=.949; median=.957) (Figure 5).

The FreeSurfer analysis produced similar, high reliability. The correlations between MRI sessions were highly significant for each regional volume and ranged from $r=.811$ to .996 (mean $r=.923$; median $r=.921$). The ICCs ranged from .801 in the posterior cingulate cortex to .995 in the lateral ventricles (mean=.919; median=.921) (Figure 5).

Discussion

Using the SRI24 atlas-based parcellation approach for analysis, the lateral ventricles and large white matter sample had exceptionally high ICCs and simple correlations as did the precuneus and lateral frontal cortical regions. Two subcortical structures, the globus pallidus and thalamus, had particularly low ICCs that improved appreciably with application of the regression correction function. With the exception of the globus pallidus (ICC corrected=.599) and the postcentral cortex (ICC corrected=.764), the corrected ICCs were all greater than .847 (.81 is considered “substantial” (Landis and Koch, 1977)).

In addition to the application of the individual ROI correction by RCF, we used two simple scalar approaches for global correction of volume differences between 1.5T and 3.0T measurements because of their ease of use and ready availability. Application of the scalar

factor based on the ratio of the mean of supratentorial volume at the two field strengths resulted in improvement of the ICCs in 7 of the 24 ROIs examined and were primarily subcortical regions. The regression of supratentorial volume at the two field strengths resulted in ICC improvement 9 ROIs. Although improvements were clearly obtained with either global correction, more consistent and greater improvements were obtained with the RCF approach. These results suggest that the 1.5T to 3.0T differences are not merely a simple volumetric scalar factor but more likely due to field strength differences in B1 inhomogeneity, tissue type conspicuity and the ability to parcellate the tissue into gray and white matter. While the RCF may be the preferred correction approach of those examined, the global scalar approach provided some improvement in measurement consistency between the field strengths, and for certain ROIs would be a preferred adjustment over no adjustment.

Similar results were obtained for the FreeSurfer analyses. Not all regional regression correction functions were the same nor were the average 1.5T vs. 3.0T mean differences necessarily in the same direction across analysis platforms, that is, between SRI24 and FreeSurfer. This is understandable given the substantial difference in anatomical designation and the difference in quantification approaches. Nonetheless, the improvement in ICCs with regression correction functions to highly acceptable levels using either method supports the contention that data collected at different field strengths can be merged.

We assessed automated morphometry reliability at 1.5T on 69 controls measured within 2 years and randomized the effect of time. Two recent reports assessed the reliability of automated morphometry at 3T over short examination intervals (Morey et al., 2010; Wonderlick et al., 2009). Morey et al. (2010) studied 23 young adults, scanned 1 hour apart on one day and again 1 hour apart 7 to 9 days later using a GE 3.0T EXCITE system with a 3D FSPGR. Of the 7 ROIs common to Morey et al. and our SRI24 results, 2 ROIs (lateral ventricles and hippocampus) had virtually the same ICC, 4 ICCs were higher (thalamus, caudate, putamen, globus pallidus) and one ICC lower (amygdala) in the Morey than our study. Wonderlick et al. (2009) scanned 5 younger (24 year old) and 6 older (64.3 year old) subjects in two identical sessions two weeks apart on a Siemens 3T TIM Trio with four variations of an MPRAGE sequences. Of the 12 ROIs common to Wonderlick et al. and our study, 8 ROIs had higher ICC, 3 ROIs had about the same ICC, and 1 ROI had lower ICCs in the Wonderlick et al. study than our study. Our 1.5T reliability results tended to have lower ICCs when we applied totally unsupervised FreeSurfer analysis.

Differences between our 1.5T reliability results and the 3T reliability of Morey et al. (2010) and Wonderlick et al. (2009) are probably due to processing differences, such as manual editing in the latter, rather than to primary differences in field strength. The principle of the current study was to determine whether a correction factor could be derived from independent observations, with the ultimate aim being the ability to use 1.5T data where 3.0T data was unavailable or vice versa. Thus, we conducted independent parcellations on the two sets of 1.5T data to ascertain reliability and the split-half 1.5T-3.0T data to determine utility. It appears Morey et al. (2010) used the FreeSurfer longitudinal analysis stream, which would likely artificially inflate the correspondence between scans by violating the assumption of independence that underlies valid scan/rescan analyses.

In an effort to examine interdependence of the measured differences across ROIs, we conducted a set of PCAs on 1.5T-3.0T difference scores. An unconstrained PCA yielded 8 factors with Eigenvalues >1, but 4 accounted for trivial variance. A follow-up PCA forced a 4-factor solution, which formed clusters with anatomical similarity: “Anterior Superior Cortex” (Factor 1 accounting for 15.6% of the variance) comprised the lateral frontal precentral, postcentral, parietal, and precuneus ROIs; “Allocortex” (Factor 2 accounting for

13.4% of the variance) comprised the anterior and posterior cingula, insula, hippocampus, corpus callosum, and Sylvian fissure ROIs; “Subcortex” (Factor 3 accounting for 11.3% of the variance) comprised the medial frontal, thalamus, caudate, putamen, centrum semiovale, and lateral and third ventricle ROIs; and “Posterior Inferior Cortex” (Factor 4 accounting for 8.1% of the variance) comprised the temporal, occipital, calcarine, and pallidal ROIs. The factor structure of the last PCA might lead to the speculation that signal intensity inhomogeneities are variable in the brain, with greater in the superior than inferior cortical regions and that variance in subcortical structure measurement could be due to accumulation of iron with age (Bartzokis et al., 2007b; Hallgren and Sourander, 1958; Pfefferbaum et al., 2010).

The 1.5T and 3.0T data that we analyzed for each subject differed in significant ways in addition to magnetic field strength. The 1.5T data were acquired in the coronal plane with a quadrature head coil, whereas the 3.0T data were acquired in the axial plane with an 8-channel phased-array head coil. Well-recognized differences in T1-weighted images at 3.0T in contrast to 1.5T include substantial B1 inhomogeneity (accounted for in part by bias correction) and the greater iron-induced relaxivity shortening at 3.0T relative to 1.5T. The latter, especially in the basal ganglia, may have contributed to the relatively poorer correspondence between the two field strengths as well as the difficulty in segmentation and parcellation of subcortical T1-weighted data; for example, the thalamus and putamen had the lowest within-1.5T scanner reliability. Although we only had available reliability estimates for 1.5T data, the very high ICCs for repeated scans on the same subjects at 1.5T suggest that the measurement methods are adequately reliable to test the field strength differences. The differences in brain regional volumes across the two field strengths were essentially equally distributed in terms of 3.0T estimates as being greater or smaller than 1.5T estimates. Thus, the field strength differences were not merely a scanner calibration or global scaling phenomenon.

The results of the split-half analysis lend support to the use of a regression-based correction function on other data sets for which it is desirable to combine data across field strengths. Different laboratories may, out of necessity, establish their own functions, which would be tailored to their acquisition parameters and quantification approaches. The current analysis was performed on two analysis platforms in an automated and unsupervised fashion. It is common practice to inspect and edit parcellation and segmentation results; for example, FreeSurfer has a large suite of editing tools, and edited data blind to field strength might have produced higher correspondence than observed here.

The intention of the regression approach presented here was to adjust for differences between data collected on different scanner platforms due to the scanning acquisition procedure itself (timing parameters, hardware, and field strength). The fact that some ROIs were larger and others smaller at 3.0T compared with 1.5T indicates that the differences were not due simply to global spatial scaling differences between the scanners. The correction depends on the existence of systematic linear differences in regional volume estimates between acquisitions, but it cannot account for changes in subjects, e.g., changes attributable to development, disease, or aging, between acquisitions. For this reason, the data used in this analysis were restricted to scans collected at 1.5T and 3.0T less than 3 weeks apart to avoid aging and disease influences on the correction factors.

In summary, the results indicate high correspondence between selected cortical, subcortical, and CSF-filled spaces that varied in a linear systematic fashion and were improved by the application of a regression-based correction function. Despite acquisition differences, the high correspondence argues well for the proposition that selected T1-weighted regional anatomical brain data can be reliably combined across 1.5T and 3.0T field strengths with the

application of an appropriate correction procedure. A similar approach could also be used for combination of data across any two scanner systems that produced linear systematic differences in structural volume estimates.

Acknowledgments

This work was supported by NIH grants AA016273, AA005965, AA017168, AG017919, EB008381. All in-house image analyses software is available from <http://nitrc.org/projects/cmtk/>. The SRI24 atlas is available from <http://nitrc.org/projects/sri24/>.

References

- Attenberger UI, Runge VM, Morelli JN, Williams J, Jackson CB, Michaely HJ. Evaluation of gadobutrol, a macrocyclic, nonionic gadolinium chelate in a brain glioma model: comparison with gadoterate meglumine and gadopentetate dimeglumine at 1.5 T, combined with an assessment of field strength dependence, specifically 1.5 versus 3 T. *J Magn Reson Imaging*. 2010; 31:549–555. [PubMed: 20187196]
- Bammer R, Hope TA, Aksoy M, Alley MT. Time-resolved 3D quantitative flow MRI of the major intracranial vessels: initial experience and comparative evaluation at 1.5T and 3.0T in combination with parallel imaging. *Magn Reson Med*. 2007; 57:127–140. [PubMed: 17195166]
- Bartzokis G, Cummings JL, Markham CH, Marmarelis PZ, Treciokas LJ, Tishler TA, Marder SR, Mintz J. MRI evaluation of brain iron in earlier- and later-onset Parkinson's disease and normal subjects. *Magn Reson Imaging*. 1999; 17:213–222. [PubMed: 10215476]
- Bartzokis G, Lu PH, Tishler TA, Fong SM, Oluwadara B, Finn JP, Huang D, Bordelon Y, Mintz J, Perlman S. Myelin Breakdown and Iron Changes in Huntington's Disease: Pathogenesis and Treatment Implications. *Neurochem Res*. 2007a; 32:1655–1664. [PubMed: 17484051]
- Bartzokis G, Sultzer D, Cummings J, Holt LE, Hance DB, Henderson VW, Mintz J. In vivo evaluation of brain iron in Alzheimer disease using magnetic resonance imaging. *Arch Gen Psychiatry*. 2000; 57:47–53. [PubMed: 10632232]
- Bartzokis G, Tishler TA, Lu PH, Villablanca P, Altshuler LL, Carter M, Huang D, Edwards N, Mintz J. Brain ferritin iron may influence age- and gender-related risks of neurodegeneration. *Neurobiology of Aging*. 2007b; 28:414–423. [PubMed: 16563566]
- Boss A, Martirosian P, Klose U, Nagele T, Claussen CD, Schick F. FAIR-TrueFISP imaging of cerebral perfusion in areas of high magnetic susceptibility differences at 1.5 and 3 Tesla. *J Magn Reson Imaging*. 2007; 25:924–931. [PubMed: 17410577]
- Chang KH, Ra DG, Han MH, Cha SH, Kim HD, Han MC. Contrast enhancement of brain tumors at different MR field strengths: comparison of 0.5 T and 2.0 T. *AJNR Am J Neuroradiol*. 1994; 15:1413–1419. discussion 1420–1413. [PubMed: 7985558]
- Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*. 1999; 9:179–194. [PubMed: 9931268]
- Dale AM, Sereno MI. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*. 1993; 5:162–176.
- Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*. 2010; 53:1–15. [PubMed: 20547229]
- Duyn JH. High-field MRI of brain iron. *Methods Mol Biol*. 2011; 711:239–249. [PubMed: 21279605]
- Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A*. 2000; 97:11050–11055. [PubMed: 10984517]
- Fischl B, Liu A, Dale AM. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans Med Imaging*. 2001; 20:70–80. [PubMed: 11293693]
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002; 33:341–355. [PubMed: 11832223]

- Fischl B, Salat DH, van der Kouwe AJ, Makris N, Segonne F, Quinn BT, Dale AM. Sequence-independent segmentation of magnetic resonance images. *Neuroimage*. 2004a; 23(Suppl 1):S69–84. [PubMed: 15501102]
- Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*. 1999; 9:195–207. [PubMed: 9931269]
- Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM. Automatically parcellating the human cerebral cortex. *Cereb Cortex*. 2004b; 14:11–22. [PubMed: 14654453]
- Fushimi Y, Miki Y, Urayama S, Okada T, Mori N, Hanakawa T, Fukuyama H, Togashi K. Gray matter-white matter contrast on spin-echo T1-weighted images at 3 T and 1.5 T: a quantitative comparison study. *Eur Radiol*. 2007; 17:2921–2925. [PubMed: 17619195]
- Goodro M, Sameti M, Patenuade B, Fein G. Age effect on subcortical structures in healthy adults. *Psychiatry Research. Neuroimaging*. 2011 in press.
- Halgren B, Sourander P. The effect of age on the non-haemin iron in the human brain. *J Neurochem*. 1958; 3:41–51. [PubMed: 13611557]
- 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. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage*. 2006; 32:180–194. [PubMed: 16651008]
- 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. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage*. 2006; 30:436–443. [PubMed: 16300968]
- 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 BC, Fischl B. 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*. 2009; 46:177–192. [PubMed: 19233293]
- Keihaninejad S, Heckemann RA, Fagiolo G, Symms MR, Hajnal JV, Hammers A. A robust method to estimate the intracranial volume across MRI field strengths (1.5T and 3T). *Neuroimage*. 2010; 50:1427–1437. [PubMed: 20114082]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
- Likar B, Viergever MA, Pernus F. Retrospective correction of MR intensity inhomogeneity by information minimization. *IEEE Transactions on Medical Imaging*. 2001; 20:1398–1410.
- Luccichenti G, Giugni E, Peran P, Cherubini A, Barba C, Bivona U, Formisano R, Sabatini U. 3 Tesla is twice as sensitive as 1.5 Tesla magnetic resonance imaging in the assessment of diffuse axonal injury in traumatic brain injury patients. *Funct Neurol*. 2010; 25:109–114. [PubMed: 20923609]
- Morey RA, Selgrade ES, Wagner HR 2nd, Huettel SA, Wang L, McCarthy G. Scanrescan reliability of subcortical brain volumes derived from automated segmentation. *Hum Brain Mapp*. 2010; 31:1751–1762. [PubMed: 20162602]
- Pfefferbaum A, Adalsteinsson E, Rohlfing T, Sullivan EV. MRI estimates of brain iron concentration in normal aging: comparison of field-dependent (FDRI) and phase (SWI) methods. *NeuroImage*. 2009; 47:493–500. [PubMed: 19442747]
- Pfefferbaum A, Adalsteinsson E, Rohlfing T, Sullivan EV. Diffusion tensor imaging of deep gray matter brain structures: Effects of age and iron concentration. *Neurobiology of Aging*. 2010; 31:482–493. [PubMed: 18513834]
- Pfefferbaum A, Chanraud S, Pitel AL, Müller-Oehring EM, Shankaranarayanan A, Alsop D, Rohlfing T, Sullivan EV. Cerebral blood flow in posterior cortical nodes of the default mode network decreases with task engagement but remains higher than in most brain regions. *Cerebral Cortex*. 2011; 21:233–244. [PubMed: 20484322]
- Pfefferbaum A, Rosenbloom MJ, Adalsteinsson E, Sullivan EV. Diffusion tensor imaging with quantitative fiber tracking in HIV infection and alcoholism comorbidity: Synergistic white matter damage. *Brain*. 2007; 130:48–64. [PubMed: 16959813]

- Pfefferbaum A, Rosenbloom MJ, Rohlfing T, Adalsteinsson E, Kemper CA, Deresinski S, Sullivan EV. Contribution of alcoholism to brain dysmorphology in HIV infection: Effects on the ventricles and corpus callosum. *NeuroImage*. 2006; 33:239–251. [PubMed: 16877010]
- Pinker K, Noebauer-Huhmann IM, Stavrou I, Hoefftberger R, Szomolanyi P, Weber M, Stadlbauer A, Grabner G, Knosp E, Trattnig S. High-field, high-resolution, susceptibility-weighted magnetic resonance imaging: improved image quality by addition of contrast agent and higher field strength in patients with brain tumors. *Neuroradiology*. 2008; 50:9–16. [PubMed: 17876570]
- Rohlfing T, Maurer CR. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Transactions on Information Technology in Biomedicine*. 2003; 7:16–25. [PubMed: 12670015]
- Rohlfing T, Maurer CR. Multi-classifier framework for atlas-based image segmentation. *Pattern Recognition Letters*. 2005; 26:2070–2079.
- Rohlfing T, Zahr NM, Sullivan EV, Pfefferbaum A. The SRI24 multi-channel atlas of normal adult human brain structure. *Human Brain Mapping*. 2010; 31:798–819. [PubMed: 20017133]
- Rosenbloom MJ, Sullivan EV, Sassoon SA, O'Reilly A, Fama R, Kemper CA, Deresinski SC, Pfefferbaum A. Alcoholism, HIV infection and their comorbidity: Factors affecting self-rated health-related quality of life. *Journal of Studies on Alcohol and Drugs*. 2007; 68:115–125. [PubMed: 17149525]
- Segonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B. A hybrid approach to the skull stripping problem in MRI. *Neuroimage*. 2004; 22:1060–1075. [PubMed: 15219578]
- Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res*. 1998; 7:301–317. [PubMed: 9803527]
- Smith S. Fast robust automated brain extraction. *Human Brain Mapping*. 2002; 17:143–155. [PubMed: 12391568]
- Srinivasan A, Goyal M, Al Azri F, Lum C. State-of-the-art imaging of acute stroke. *Radiographics*. 2006; 26(Suppl 1):S75–95. [PubMed: 17050521]
- Stankiewicz JM, Glanz BI, Healy BC, Arora A, Neema M, Benedict RH, Guss ZD, Tauhid S, Buckle GJ, Houtchens MK, Khoury SJ, Weiner HL, Guttmann CR, Bakshi R. Brain MRI lesion load at 1.5T and 3T versus clinical status in multiple sclerosis. *J Neuroimaging*. 2011; 21:e50–56. [PubMed: 19888926]
- Wonderlick JS, Ziegler DA, Hosseini-Varnamkhasi P, Locascio JJ, Bakkour A, van der Kouwe A, Triantafyllou C, Corkin S, Dickerson BC. Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging. *Neuroimage*. 2009; 44:1324–1333. [PubMed: 19038349]
- Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Transactions Medical Imaging*. 2001; 20:45–57.
- Zhu YC, Dufouil C, Tzourio C, Chabriat H. Silent brain infarcts: a review of MRI diagnostic criteria. *Stroke*. 2011; 42:1140–1145. [PubMed: 21393597]
- Zijlmans M, de Kort GA, Witkamp TD, Huiskamp GM, Seppenwoolde JH, van Huffelen AC, Leijten FS. 3T versus 1.5T phased-array MRI in the presurgical work-up of patients with partial epilepsy of uncertain focus. *J Magn Reson Imaging*. 2009; 30:256–262. [PubMed: 19629993]

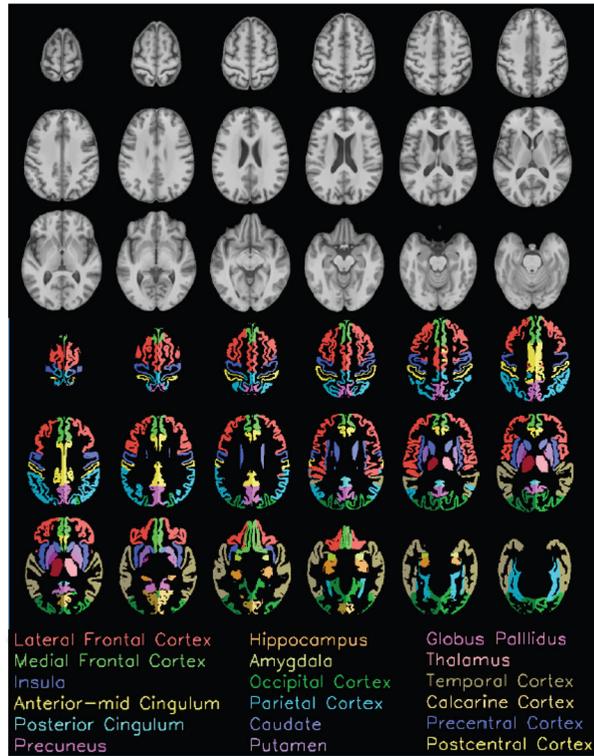


Figure 1.

Top panel in gray scale: Axial slices from the SRI24 atlas of the superior (top left) to inferior (bottom right) reaches of the brain. Bottom panel in color: 24 cortical and subcortical regions overlaid on the SRI24 atlas and color-coded by structure name.

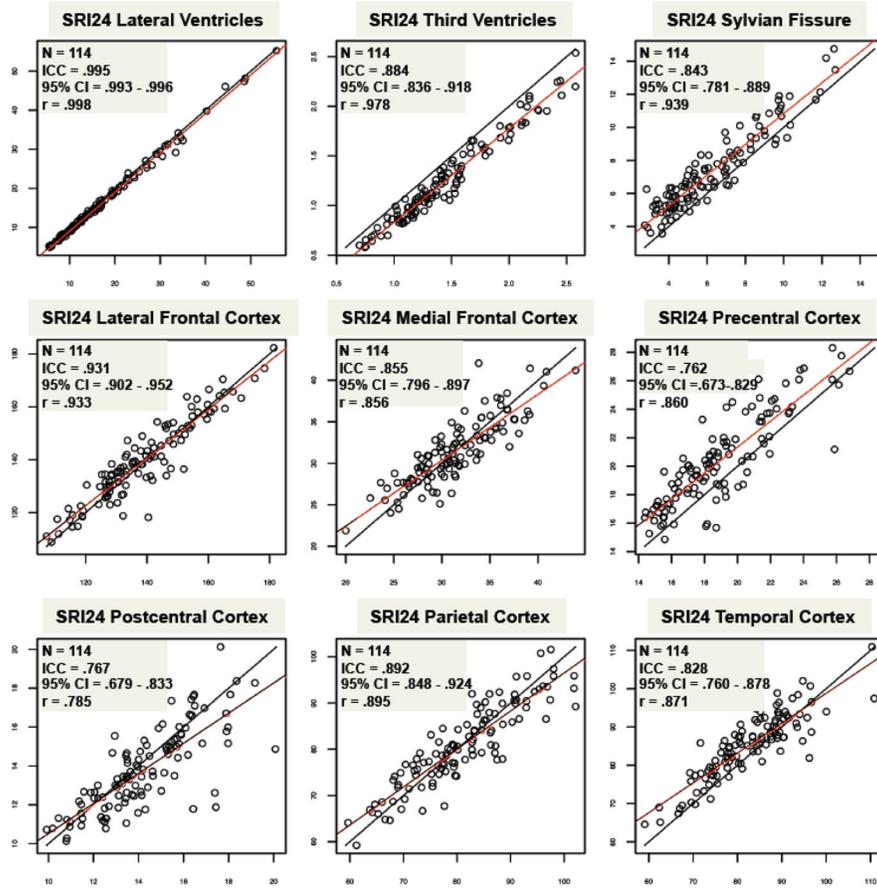


Fig. 2a

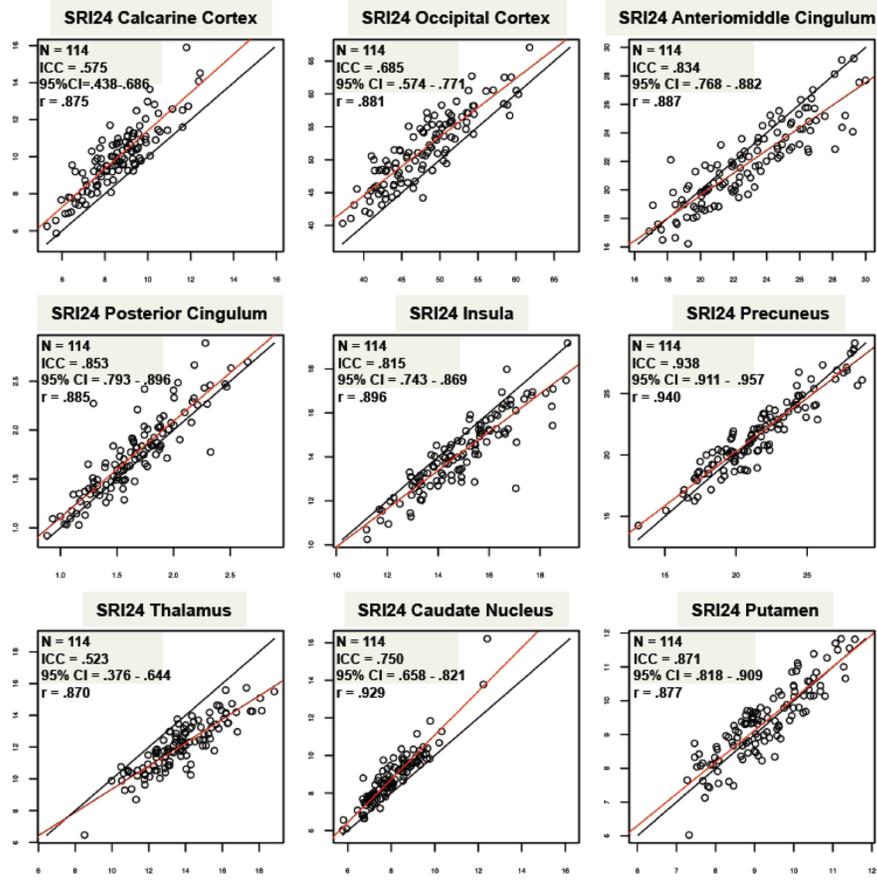


Fig. 2b

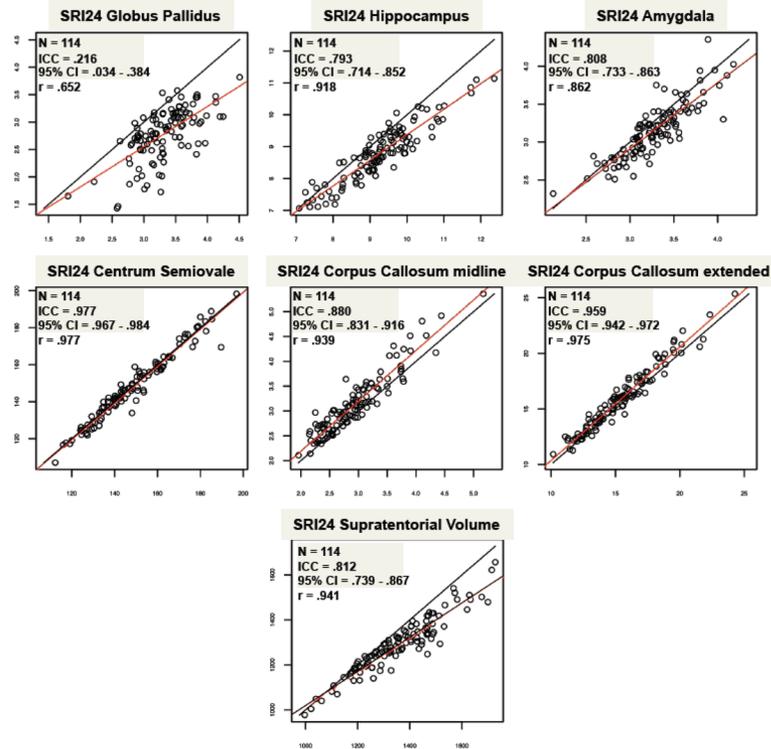


Fig. 2c

Figure 2.
Figure 2a–c. ICCs for each SRI24 atlas-derived regional volume pair of 1.5T (x-axis) and 3.0T (y-axis) data from 114 adults scanned at each field strength within 3 weeks.

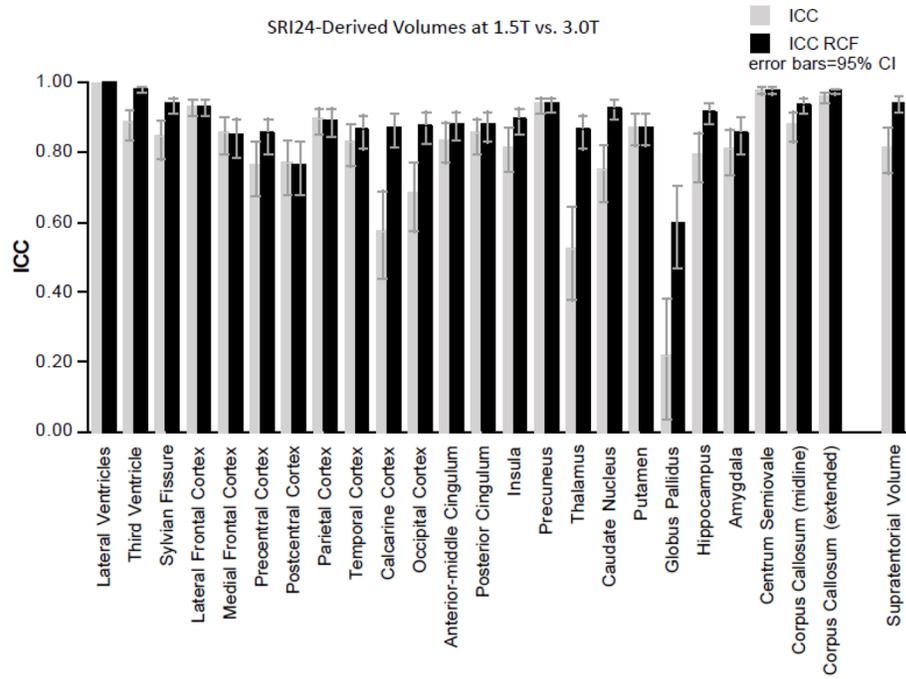


Figure 3. ICCs before and after application of regression correction function (RCF) for each SRI24 atlas-derived regional volume pair of 1.5T and 3.0T data from 114 adults scanned at each field strength within 3 weeks.

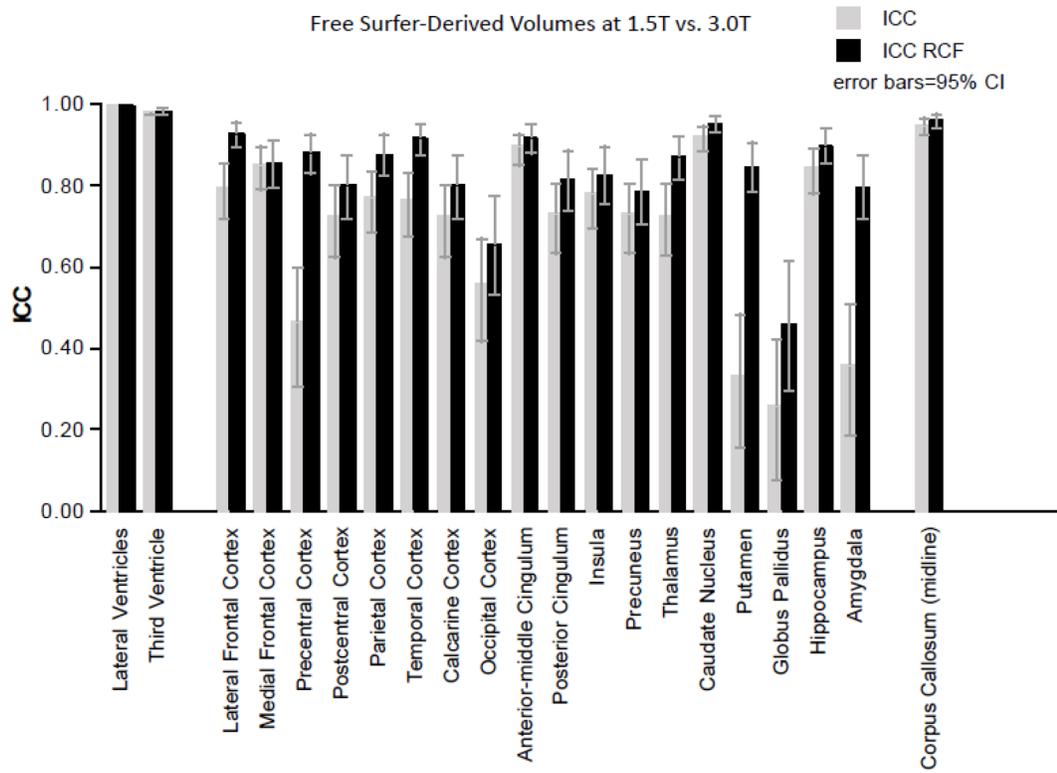


Figure 4. ICCs before and after application of regression correction function (RCF) for each FreeSurfer-derived regional volume pair of 1.5T and 3.0T data from 114 adults scanned at each field strength within 3 weeks.

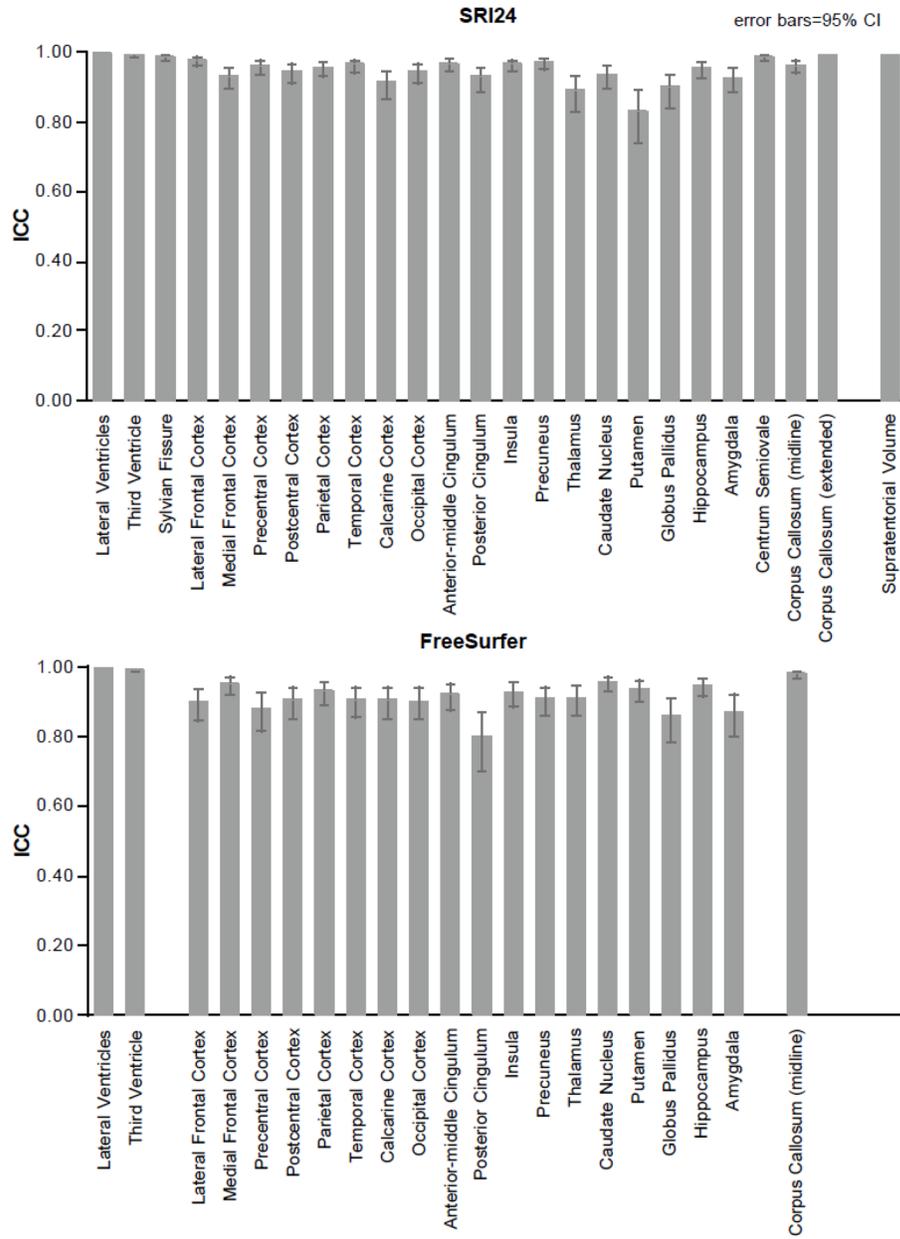


Figure 5. ICCs for each SRI24 atlas-derived (top) and FreeSurfer-derived (bottom) regional volume pair of 1.5T data collected twice in 69 healthy adults.

Table 1

1.5T-3.0T comparisons of SRI24-derived regional brain volumes (mm³) in 114 subjects

ROI	1.5T mean volume (SD)	3.0T mean volume (SD)	ICC	95% CI	r	1.5T/3.0T	slope	intercept	ICCRCF	95% CI
Lateral Ventricles	17.42 (10.00)	16.66 (9.95)	.995	(.993-.996)	.998	1.045	.993	-6.29	.998	(.997-.998)
Third Ventricle	1.45 (.44)	1.26 (.43)	.884	(.836-.918)	.978	1.154	.944	-1.13	.978	(.968-.985)
Sylvian Fissure	6.23 (2.41)	7.33 (2.38)	.843	(.781-.889)	.939	.850	.931	1.529	.938	(.911-.957)
Lateral Frontal Cortex	139.28 (15.68)	140.12 (15.29)	.931	(.902-.952)	.933	.994	.910	13.420	.931	(.902-.952)
Medial Frontal Cortex	31.31 (4.27)	31.39 (3.95)	.855	(.796-.897)	.856	.997	.793	6.553	.847	(.786-.892)
Precentral Cortex	18.87 (2.96)	20.31 (3.14)	.762	(.673-.829)	.860	.929	.912	3.099	.852	(.792-.895)
Postcentral Cortex	14.22 (2.09)	13.79 (2.06)	.767	(.679-.833)	.785	1.032	.774	2.771	.764	(.676-.831)
Parietal Cortex	81.13 (9.62)	80.95 (8.83)	.892	(.848-.924)	.895	1.002	.822	14.288	.891	(.845-.923)
Temporal Cortex	82.69 (9.23)	85.25 (8.27)	.828	(.760-.878)	.871	.970	.780	20.768	.864	(.809-.904)
Calcarine Cortex	8.61 (1.46)	9.97 (1.72)	.575	(.438-.686)	.875	.863	1.031	1.095	.868	(.814-.907)
Occipital Cortex	47.99 (5.36)	51.68 (5.42)	.685	(.574.771)	.881	.929	.892	8.884	.876	(.825-.912)
Anterior-middle Cingulum	22.88 (3.19)	21.89 (2.85)	.834	(.768-.882)	.887	1.045	.793	3.750	.881	(.833-.917)
Posterior Cingulum	1.65 (.37)	1.75 (.41)	.853	(.793-.896)	.885	.945	.989	.116	.880	(.831-.915)
Insula	14.89 (1.73)	14.17 (1.69)	.815	(.743-.869)	.896	1.051	.872	1.175	.892	(.847-.924)
Precuneus	21.58 (3.27)	21.64 (3.06)	.938	(.911-.957)	.940	.997	.878	2.687	.938	(.912-.957)
Thalamus	13.75 (1.90)	12.10 (1.60)	.523	(.376-.644)	.870	1.136	.734	2.014	.862	(.807-.903)
Caudate Nucleus	8.05 (1.13)	8.81 (1.41)	.750	(.658.821)	.929	.913	1.164	-5.57	.927	(.896-.949)
Putamen	9.24 (1.03)	9.35 (1.10)	.871	(.818-.909)	.877	.988	.942	.642	.870	(.818-.909)
Globus Pallidus	3.31 (.42)	2.78 (.47)	.216	(.034-.384)	.652	1.191	.734	.349	.599	(.468-.705)
Hippocampus	9.29 (1.01)	3.80 (.88)	.793	(.714-.852)	.918	1.055	.803	1.346	.915	(.880-.941)
Amygdala	3.28 (.35)	3.15 (.36)	.808	(.733-.863)	.862	1.039	.876	.280	.854	(.795-.897)
Centrum Semiovale	148.40 (17.98)	147.76 (18.29)	.977	(.967-.984)	.977	1.004	.994	.203	.977	(.967-.984)
Corpus Callosum (midline)	2.92 (.56)	3.12 (.61)	.880	(.831.916)	.939	.936	1.015	.157	.937	(.910-.956)
Corpus Callosum (extended)	15.41 (2.67)	15.87 (2.80)	.959	(.942-.972)	.975	.971	1.023	.112	.975	(.963-.982)
Supratentorial Volume	1353.69 (148.24)	1287.76 (119.70)	.812	(.739-.867)	.941	1.051	.760	258.697	.940	(.915-.958)

ICCRCF = ICCs after application of regression correction function (RCF)

CI =confidence interval

1.5T/3.0T is the (mean 1.5T)/(mean 3.0T)

Table 2

1.5T-3.0T comparisons of SRI24-derived RCF ICCs after correction in random split samples: 57 subjects/
sample

ROI	a-b		b-a	
	RCF ICC	95% CI	RCF ICC	95% CI
Lateral ventricles	.998	(.997-.999)	.997	(.995-.998)
Third Ventricle	.976	(.960-.986)	.977	(.962-.986)
Sylvian Fissure	.937	(.896-.963)	.934	(.891-.961)
Lateral Frontal Cortex	.909	(.851-.945)	.941	(.902-.965)
Medial Frontal Cortex	.826	(.725-.894)	.860	(.774-.915)
Precentral Cortex	.831	(.730-.897)	.859	(.772-.914)
Postcentral Cortex	.686	(.520-.802)	.819	(.712-.889)
Parietal Cortex	.860	(.774-.915)	.919	(.867-.951)
Temporal Cortex	.841	(.744-.903)	.869	(.788-.921)
Calcarine Cortex	.847	(.754-.907)	.875	(.797-.924)
Occipital Cortex	.880	(.805-.927)	.872	(.793-.922)
Anterior-middle Cingulum	.886	(.814-.931)	.871	(.791-.922)
Posterior Cingulum	.851	(.761-.909)	.903	(.842-.942)
Insula	.894	(.826-.936)	.885	(.813-.931)
Precuneus	.935	(.892-.961)	.935	(.893-.961)
Thalamus	.825	(.721-.893)	.809	(.696-.882)
Caudate Nucleus	.927	(.880-.956)	.920	(.868-.952)
putamen	.786	(.663-.868)	.815	(.706-.887)
Globus Pallidus	.585	(.386-.732)	.454	(.223-.637)
Hippocampus	.867	(.785-.919)	.936	(.893-.961)
Amygdala	.814	(.704-.886)	.887	(.815-.932)
Centrum Semiovale	.972	(.953-.983)	.980	(.967-.988)
Corpus Callosum (midline)	.929	(.883-.958)	.926	(.878-.956)
Corpus Callosum (extended)	.969	(.948-.981)	.975	(.958-.985)
Supratentorial Volume	.926	(.877-.955)	.943	(.905-.966)

CI=confidence interval

a-b=regression correction factor of sample, a applied to sample b

b-a=regression correction factor of sample b applied to sample a

RCF ICC = ICCs after application of regression correction function (RCF)

Table 3
1.5T to 3.0T comparisons of FreeSurfer-derived regional brain volumes in 114 subjects

ROI	ICC	95% CI	r	1.5T/3.0T	Slope	Intercept	ICCRCF	95% CI
Lateral Ventricles	.997	(.996-.998)	.998	.979	1.006	.274	.998	(.997-.998)
Third Ventricle	.982	(.974-.987)	.983	1.017	1.004	-0.22	.983	(.975-.988)
Sylvian Fissure	•	•	•	•	•	•	•	•
Lateral Frontal Cortex	.796	(.718-.855)	.930	.944	.954	11.310	.928	(.898-.950)
Medial Frontal Cortex	.850	(.790-.894)	.861	.980	.900	4.705	.853	(.794-.896)
Precentral Cortex	.466	(.310-.598)	.885	.878	.989	3.462	.879	(.830-.915)
Postcentral Cortex	.725	(.625-.802)	.814	.944	.890	2.942	.798	(.720-.856)
Parietal Cortex	.772	(.686-.837)	.881	.948	.988	7.357	.875	(.824-.912)
Temporal Cortex	.763	(.675-.830)	.917	.935	.983	6.275	.914	(.878-.940)
Calcarine Cortex	.725	(.625-.802)	.814	.944	.890	2.942	.798	(.720-.856)
Occipital Cortex	.557	(.416-.671)	.695	.931	.908	5.542	.654	(.535-.747)
Anterior-middle Cingulum	.894	(.850-.926)	.919	.968	.965	.983	.916	(.881-.941)
Posterior Cingulum	.731	(.633-.806)	.826	1.053	.759	2.641	.813	(.740-.867)
Insula	.779	(.695-.842)	.837	.953	.960	1.629	.826	(.757-.876)
Precuneus	.731	(.633-.806)	.804	.941	.962	.497	.787	(.706-.848)
Thalamus	.727	(.627-.803)	.877	1.080	.850	1.056	.870	(.817-.908)
Caudate Nucleus	.920	(.886-.944)	.951	.965	.972	.439	.950	(.929-.965)
Putamen	.332	(.159-.485)	.855	.865	1.071	.827	.846	(.785-.891)
Globus Pallidus	.258	(.079-.421)	.542	1.118	.586	1.045	.457	(.299-.591)
Hippocampus	.844	(.781-.889)	.901	.960	.975	.562	.897	(.854-.927)
Amygdala	.358	(.187-.508)	.812	.865	1.135	.066	.796	(.718-.855)
Centrum Semiovale	•	•	•	•	•	•	•	•
Corpus Callosum (midline)	.948	(.926-.964)	.960	.976	.995	.089	.959	(.941-.972)
Corpus Callosum (extended)	•	•	•	•	•	•	•	•
Supratentorial Volume	•	•	•	•	•	•	•	•

ICCRCF = ICCs after application of regression correction function (RCF)

CI =confidence interval

1.5T/3.0T is the (mean 1.5T)/(mean 3.0T)