PHANTOM-BASED MRI CORRECTIONS AND POWER TO TRACK BRAIN CHANGE

Christopher R. K. Ching[1], Xue Hua[1], Chadwick Ward[2], Jeff Gunter[2], Matt A. Bernstein[2], Clifford R. Jack Jr[2], Michael W. Weiner[3,4,5], Paul M. Thompson[1,6]

1. Laboratory of Neuro Imaging, Dept. of Neurology, UCLA School of Medicine, CA, USA
2. Mayo Clinic and University of Minnesota, Rochester, MN, USA
3. Dept. of Radiology, UCSF, CA, USA
4. Dept. Medicine and Psychiatry, UCSF, CA, USA
5. Dept. Veterans Affairs Medical Center, San Francisco, CA, USA
6. Dept. of Neurology and Psychiatry, UCLA School of Medicine, CA, USA

ABSTRACT

The Alzheimer's Disease Neuroimaging Initiative (ADNI) employed phantom-based scaling of T1-weighted MP-RAGE brain images to improve spatial calibration of scans and longitudinal stability across study sites. Early in ADNI, errors in phantom based scaling were identified due to incorrect acquisition protocols or replacement of phantoms. Images initially made available with questionable scaling factors (scaled) were reprocessed (scaled_2) and made available to the scientific community. As many studies were conducted using images with sub-optimal scaling, we aimed to determine if the power to detect brain change differed between scaled and scaled_2 images. Using tensor-based morphometry, we found a high degree of correlation and no significant difference between scaled and scaled_2 images. We detected no significant differences between effect sizes derived from scaled and scaled_2 images. Our findings support the ADNI MRI core assessment that analyses carried out with either scaled or scaled_2 images may not offer a substantial difference in power.

Index Terms— Magnetic Resonance Imaging, Phantoms, Tensor-Based Morphometry, Brain Modeling

1. INTRODUCTION

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a multi-site longitudinal study with the goal of developing and evaluating reliable imaging biomarkers to track and predict the progression of Alzheimer's disease (AD). ADNI recruited 842 subjects at 58 North American sites using standardized protocols. The ADNI data are freely available online [1]. Due to imperfections in the spatial calibrations of scanners, images collected at different sites, or at the same site over time, may differ slightly in their geometric calibration. To compensate for differences in spatial calibration across time and across ADNI scanning sites, phantom scans were used to correct for linear and nonlinear spatial distortions, as well as variations in signal-to-noise ratio (SNR) and image contrast across sites. In addition to adjusting scans using phantom-based geometric corrections**,** phantom scanning can allow investigators to quickly identify sites with incorrect spatial calibration, or temporal drift beyond a range deemed acceptable. Early on in the ADNI project, several well-documented errors (see Methods) in phantom scaling were identified that may help in designing future studies [2]**.** When those errors were identified, an effort was made to quickly correct the scaling for the erroneous images. Corrected images were re-uploaded and made available to the scientific community [3]. As several analyses had already been published using images with known errors due to phantom based scaling, here we set out to estimate the impact of phantom scaling errors on one popular type of longitudinal analysis of brain change - tensor-based morphometry (TBM). We used the most relevant subset of the ADNI data, from all subjects who have images with both sets of scaling factors (scaled vs. scaled_2 images). We examined the correlation between changes computed from scaled and scaled_2 images in a TBM analysis. We also determined whether there was any bias depending on the scaling factor (over- or under-estimate of change), and tested for any difference in effect sizes calculated from scaled and scaled_2 images. In short, we wanted to see if the correction improved the power to detect brain change, or whether it made no significant difference, at least for one popular type of analysis of brain change.

2. METHODS

ADNI was launched in 2003 as a 5-year study collecting (among other biomarkers) brain MRI data from over 842 adults (aged 55-90). Baseline data was collected from 200 patients with Alzheimer's disease (AD), 410 with mild cognitive impairment (MCI) and 232 cognitively normal elderly controls (CN). All subjects were scanned every 6 months to help evaluate the statistical power of methods to detect brain change.

2.1 Phantom Scaling

The ADNI phantom was used to compute a geometric scaling transform to correct for scanner and session-specific calibration errors. Problems in phantom scaling of T1-weighted MRI images were identified in two situations:

1. Some ADNI phantoms had to be replaced due to manufacturing defects or on-site damage; this resulted in unreliable scaling in the A/P direction.
2. Eleven 1.5T MRI scanners initially received an incorrect protocol parameter in which autoshim was disabled, resulting in unreliable scaling in the S/I dimension.

In October 2008, a new set of scans termed "scaled_2 scans" were created to set scaling to the value of 1.0 on the axis for which the accuracy of phantom based scaling was in question. These new scaled_2 scans were made accessible online [1]. The details of phantom based scaling in ADNI are well documented in a prior publication [4].

2.2 Image acquisition

We downloaded 1.5 Tesla T1-weighted MP-RAGE images for subjects that had both a scaled and scaled_2 scan available. **Table 1** summarizes the scans available on the date of download (7/14/2011). All subjects underwent clinical and cognitive assessment when the scans were acquired. The ADNI protocol is available online [1].

	6 Month	12 Month	18 Month	24 Month
AD	**27**	**15**	**0**	**0**
MCI	**72**	**36**	**13**	**3**
CN	**44**	**23**	**0**	**3**
Total	**143**	**74**	**13**	**6**

Table 1: Total sample sizes at each time point separated by diagnosis. AD: Alzheimer's disease; MCI: Mild cognitive impairment; CN: Cognitively Normal Control

2.3 Image Analysis: Tensor Based Morphometry (TBM)

TBM is an automated technique used to identify regional structural differences between MRI images, as well as brain changes over time. To estimate brain change in each subject, follow-up scans were linearly registered with a 9 parameter (9P) linear transformation to their corresponding screening scan (SC) and both scans were then aligned to the standard ICBM space using the same 9P registration derived from spatial alignment of the SC to the ICBM [5]. Second, a non-linear inverse-consistent elastic intensity-based registration algorithm was utilized to assess volumetric tissue differences at a voxel-wise level, also known as the determinant of the Jacobian matrix [6]. TBM processing is further detailed in [7]. To enforce inverse-consistency, we used the algorithm from [6] for both cross-sectional and longitudinal analyses.

2.4 Numerical summaries

Numerical summaries were derived from 3D rate-of-atrophy maps to summarize the amount of cumulative atrophy over 6, 12, 18 and 24 months, in a statistically-defined ROI (stat-ROI, computed from a non-overlapping AD training sample, $p<0.00001$), and anatomically-defined ROIs (temporal lobe and temporal lobe gray matter). Methods to derive these numerical summaries are described in [7].

2.5 Statistics

Scatter plots were generated to compare scaled and scaled_2 numerical summaries at each time point. Correlation coefficients (R^2) and *p*-values from paired two-sample *t*-tests were calculated. Power analysis was conducted at 6 and 12 months, but not at later time points due to the limited number of subjects. As defined by the ADNI Biostatistics Core, the sample size was estimated that would be required to detect a 25% reduction in the mean annual rate of atrophy with 80% power using a two-sided test with a standard significance level ($\alpha=0.05$) for a hypothetical two-arm study. These sample sizes are referred to as "n80's", and are computed as in [8]. Confidence intervals (95%) for each n80 estimate were computed from 10,000 bootstrapped samples [7].

3. RESULTS

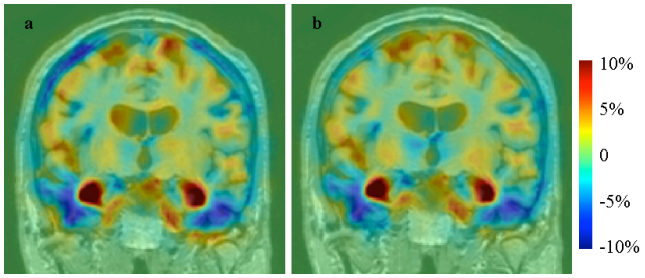

Figure 1: Visual comparison of whole-brain Jacobian maps for a 67 year old male AD subject at a 12-month follow up derived from scaled (a) and scaled_2 (b) scans. In the same coronal slice, the Jacobian maps show visually similar levels of temporal lobe atrophy (*blue*) and ventricular expansion (*red*), over this 1-year follow-up interval, although minor differences are visible in some CSF regions outside the brain (*top left*).

Figure 2 compares TBM numerical summary measures from scaled and scaled_2 images. **Table 2** shows the correlation coefficients for the comparisons (all greater than 0.71). Correlations were greater when the changes were larger (24 months).

	6 Month	12 Month	18 Month	24 Month
Stat-ROI (p<.00001)	**0.42**	**0.49**	**0.27**	**0.81**
Temporal lobe ROI	**0.68**	**0.72**	**0.11**	**0.90**
Temporal GM ROI	**0.78**	**0.77**	**0.15**	**0.81**

Table 2: Squared correlation coefficients (R^2 values) for linear regressions fitted to compare numerical summaries (% cumulative atrophy) from scaled vs. scaled_2 images, for the entire ADNI cohort (AD+MCI+CN) at each follow up time-point.

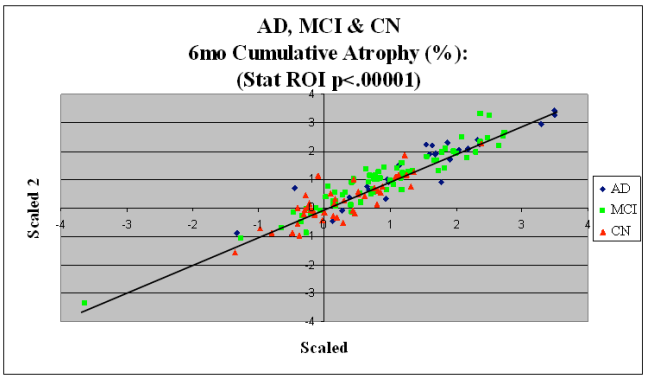

Figure 2: Plot comparing 6 month scaled and scaled_2 numerical summaries taken from the Stat-ROI (with threshold p<0.00001) across all diagnoses. Each point represents the scaled (x axis) and scaled_2 (y axis) numerical summary for a single subject. Individual subjects are color-coded by diagnosis. Measures are highly but not perfectly correlated (R^2=0.88); there is no detectable bias in the amount of change detected.

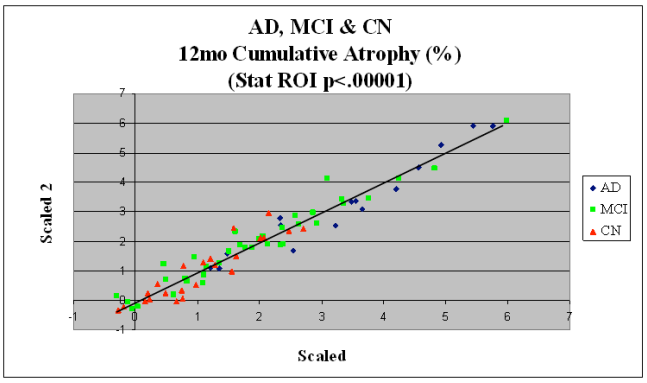

Figure 3: Plot comparing 12-month scaled and scaled_2 numerical summaries taken from the stat-ROI (with threshold p<0.00001) across all diagnoses. Measures are highly but not perfectly correlated (R^2=0.94).

Paired two-sample t-tests were calculated to detect any differences between scaled and scaled_2 numerical summaries at each time point. As summarized in **Table 3**, no differences were detected (all $p > 0.05$).

	6 Month	12 Month	18 Month	24 Month
Stat-ROI (p<.00001)	**0.88**	**0.94**	**0.87**	**0.98**
Temporal lobe ROI	**0.73**	**0.78**	**0.86**	**0.91**
Temporal GM ROI	**0.72**	**0.78**	**0.88**	**0.93**

Table 3: P-values from paired two-sample t-tests comparing numerical summaries (% cumulative atrophy) from scaled and scaled_2 images for the combined ADNI group (AD+MCI+CN) at each follow up time-point. Despite the large sample, no difference was detected between scaled and scaled_2 scans (all p>0.05).

At 6 months, where enough subjects were available, correlation coefficients and p-values from paired two-sample t-tests for each diagnostic group were calculated for each diagnostic group (AD, MCI and CN). We found no significant difference between scaled and scaled_2 images within diagnostic groups, all R^2>0.69.

As the available sample size was lower after 12 months, we only computed n80's for 6 and 12 month follow-up intervals. Twelve-month results are shown in **Table 4** (*last page*). Due to space limits, 6-month results are not shown, but show a similar pattern. Effect sizes and n80's were not detectably different for scaled versus scaled_2 images.

4. CONCLUSION

There are two main findings for this study. First, we found a high degree of correlation and no significant difference between scaled and scaled_2 numerical summaries derived from TBM analysis. For groups with adequate sample size (6-month), we found no difference and a high correlation between scaled and scaled_2 images even within diagnostic groups. Second, n80 measures for all three numerical summaries were similar when derived from scaled versus scaled_2 images.

One reason for such close correlation between scaled and scaled_2 images in TBM analysis may be the use of 9 parameter scaling as an image pre-processing step. Nine parameter linear registration allows independent scaling in x-, y-, and z- dimension and it has been shown to outperform 6P registration to correct for scanner voxel size variations in large studies involving multiple sites, scanners, and acquisition sequences [7,9]. Furthermore, 9P registration has been shown to produce similar levels of scaling correction compared to phantom-based image correction methods in ADNI [3]. In the case of TBM

analyses, it may be unnecessary to re-do studies conducted with scaled data that has been 9P registered, as we detected no significant improvement in power (or any consistent difference at all) when scaled_2 images were used. Because we did not test for a difference in scaled and scaled_2 images using 6P registrations we cannot say whether past experiments using 6P scaled data should be repeated. Most modern image processing pipelines now include a 9- or 12-parameter adjustment for brain scale, so phantom-based scaling were in fact abandoned for ADNI2, the follow-on project from ADNI.

Our finding of no significant difference between scaled and scaled_2 images in terms of TBM analysis agrees with suggestions made by the ADNI MRI core [10]. Moving forward, the MRI core has suggested that scaled_2 images be used for subjects that have both scaled and scaled_2 images uploaded to the ADNI database. Based on **Table 4**, our power analysis suggests that, at least in the case of TBM, scaled images may sometimes appear to offer more power to detect change (smaller n80's) depending on the ROI and diagnostic group, although the confidence limits overlap.

5. REFERENCES

[1] http://adni.loni.ucla.edu/

[2] Gunter J. et al. Validation testing of the MRI calibration phantom for the Alzheimer's Disease Neuroimaging Initiative Study. ISMRM 14th Scientific Meeting and Exhibition (2006).

[3] Clarkson, M.J. et al. Comparison of phantom and registration scaling corrections using the ADNI cohort. NeuroImage 47, 1506–1513. 2009

[4] Gunter J. et al. Measurement of MRI scanner performance with the ADNI phantom. Med Phys. 36(6): 2193-205 (2009)

[5] Mazziotta J. et al. A probabilistic atlas and reference system for the human brain. International Consortium for Brain Mapping (ICBM). Philos. Trans. R. Soc. Lond., B Biol. Sci. 356, 1293-1322 (2001)

[6] Leow A. et al. Inverse consistent Mapping in 3D Deformable Image Registration: Its Construction and Statistical Properties. Information Processing in Medical Imaging. Springer Berlin, Glenwood Springs, Colorado, USA 493-503 (2005)

[7] Hua X. et al. Accurate measurement of brain changes in longitudinal MRI scans using tensor-based morphometry. Neuroimage 57, 5-14 (2011)

[8] Rosner, B. Fundamentals of Biostatistics. PWS-Kent Publishing Company, Boston. (1990)

[9] Paling, S.M. et al. The application of serial MRI analysis techniques to the study of cerebral atrophy in late-onset dementia. Med Image Anal. 8(1):69-79 (2004)

[10] http://adni.loni.ucla.edu/2008/10/scaled-2-uploads/

AD	Scaled_2 Stat ROI p<.00001	Scaled Stat ROI p<.00001	Scaled_2 Temporal ROI	Scaled Temporal ROI	Scaled_2 Temporal GM ROI	Scaled Temporal GM ROI
% Tissue atrophy	3.2287	3.3401	1.2595	1.3209	1.7603	1.8487
Std	1.5988	1.4591	0.6593	0.5476	0.8101	0.6981
n80	62 [36, 116]	48 [26, 90]	69 [40, 123]	44 [20, 113]	54 [31, 97]	36 [14, 96]
MCI	Scaled_2 Stat ROI p<.00001	Scaled Stat ROI p<.00001	Scaled_2 Temporal ROI	Scaled Temporal ROI	Scaled_2 Temporal GM ROI	Scaled Temporal GM ROI
% Tissue atrophy	1.9499	1.9345	0.9083	0.8958	1.1402	1.1152
Std	1.4354	1.4254	0.7891	0.7871	1.0579	1.0528
n80	136 [83, 241]	136 [85, 244]	189 [99, 527]	194 [106, 458]	216 [109, 601]	224 [120, 523]
CN	Scaled_2 Stat ROI p<.00001	Scaled Stat ROI p<.00001	Scaled_2 Temporal ROI	Scaled Temporal ROI	Scaled_2 Temporal GM ROI	Scaled Temporal GM ROI
% Tissue atrophy	1.0198	1.0699	0.4069	0.4334	0.5444	0.5773
Std	1.0007	0.8417	0.5008	0.4670	0.6742	0.6003
n80	242 [124, 550]	155 [85, 359]	380 [155, 4007]	291 [113, 1483]	385 [144, 3550]	271 [98, 1481]

Table 4: 12 month effect sizes for brain changes, split by diagnostic group. Mean % tissue atrophy over the interval, standard deviation (Std) of the % atrophy, and n80 [confidence interval] are provided for each of the three numerical summaries for both scaled and scaled_2 images by diagnostic group. As expected, the change in the statistical ROI is higher than that in the temporal lobe overall, as it focuses on voxels expected to change the most.