Supplemental Information

for

Normal-Appearing Cerebral White Matter in Healthy Adults: Mean Change over Two Years and Individual Differences in Change

Andrew R. Bender and Naftali Raz

Institute of Gerontology & Department of Psychology, Wayne State University

1. Methods

1.1 Participants

Only 104 out of 134 participants who completed longitudinal assessment at both occasions, were tested on the 4T scanner; the remaining 30 participants were scanned on a 3T magnet that replaced it, a change that was beyond our control. Comparison of a small sample scanned back-to-back on 4T and 3T magnets established significant differences among the collected data sets, which precluded combined analyses (Raz, unpublished data). Thus, only participants scanned on the 4T scanner were eligible for this study

Initially, 219 participants (67.6% women) completed baseline assessments, and of those, 134 returned for follow up. The sample of 134 participants who completed both waves of the study did not differ from the group of 85 drop-outs with regard to the sex composition (χ^2 = 2.60, $p > 0.10$), frequency of diagnosed and treated hypertension (χ^2 = 1.20, $p > 0.10$), years of formal education ($t[217] = 0.28$, $p > 0.10$, systolic ($t[216] = 0.33$, $p > 0.10$) and diastolic ($t[216] = 1.28$, $p > 0.10$) blood pressure, self-reported leisure-time physical activity (χ^2 = 0.88, $p > 0.10$), and frequency thereof ($t[217] = 1.65$, $p = 0.100$), as well as the proportion of smokers (χ^2 = 1.00, $p > 0.10$). However, the participants who remained in the study, were older (mean age = 53.37, SD = 13.82 years vs. 48.49, SD = 17.32 years, $t[217] = -2.30$, $p < .05$) and had slightly higher mean MMSE scores (mean MMSE = 28.98, SD = 0.99 vs. 28.60, SD = 1.18; $t[217] = -2.55$, $p < .05$).

1.2 DTI Processing

1.2.1 DTI Processing Pipeline. All DTI data were processed using a custom pipeline, implemented in the FMRIB Software Library (FSL) v5.0.2 (Analysis Group, FMRIB, Oxford, UK). The pipeline was written, tested and executed on an Apple Mac Pro workstation running Mac OS 10.7 (Apple, Inc., Cupertino, CA). The pipeline was designed with multiple overarching goals: 1) Assess the same anatomical regions from two occasions of measurement; 2) Minimize

undue interpolation resulting from spatial transformation by sampling from native space images; and 3) Optimize the signal from normal appearing white matter by excluding cerebral-spinal fluid (CSF) and white matter hyperintensities (WMH) visible on T2-weighted images. This was achieved by pre-registration of data from the two measurement occasions for each participant, using the tract-based spatial statistics (TBSS; Smith et al., 2006) processing pipeline to reduce FA images to a group-wise, mean white matter (WM) 'skeleton,' one voxel wide, using the T2-weighted b_0 images to mask CSF and visible WMH from native space images, and deprojection and sampling of the mean WM skeleton and atlas-derived skeletonized regions of interest (ROIs) from the CSF/WMH-masked native space data. In addition, data are thresholded at multiple points to minimize noise in the final signal. Although these may be considered redundant steps, such a conservative approach should only serve to maximize signal to noise ratio. A complete description of all processing steps follows.

1.2.2 Post-acquisition. Images were averaged across acquisitions on the scanner console and saved in DICOM format. We converted image data from DICOM format into NIFTI-1 (.nii) using MRIConvert 2.0 (Jolinda Smith, University of Oregon, Lewis Center for Neuroimaging [http://lcni.uoregon.edu/~jolinda/mriconvert]), which also extracted and rotated the b-vector matrix for each participant by multiplying the original b-vector gradient matrices by the inverse of the individual rotation matrices (i.e., the patient image orientation matrix).

1.2.3 DTI Pre-processing. Following extraction of the first (b0) volume from each of the two paired (Time 1, Time 2) 4D files, the FMRIB's Linear Image Registration Tool (FLIRT; Jenkinson and Smith, 2001) was used to linearly register the b0 images from each occasion using a 6-degrees-of-freedom (df) rigid registration with tri-linear interpolation, saving the transformation matrices for subsequent use. Following registration, the script invoked the 'avscale' command to calculate the matrices needed to transform both images into the intermediate or 'halfway' space between them. The 'convert_xfm' command extracted the halfway-forward and halfway-backward matrices and inverted copies of the two matrices for

later use.

We used the brain extraction tool (BET; Smith, 2002) in FSL on the paired, untransformed b_0 images and the native 4D .nii file to produce brain masks and remove non-brain tissue; BET parameters included the –m and –R flags for binary mask creation and robust repeated estimation of the brain center, respectively. The b_0 images, stripped of non-brain tissue, were subsequently eroded by one voxel. The 4D brain masks were used to fit the tensor using the dtifit function, as well as participant specific b-vectors, a common b-value of 800 mm^2/s, and the 'save_tensor' function to retain the DTI components for later use.

Next, the 'vecreg' command rotated the saved tensor components using the halfway transformation matrices (i.e., Time 1 – halfway backward, Time 2 – halfway forward) that were generated on the previous step using convert_xfm. The fslmaths 'tensor-decomp' function refit the tensor data for both occasions in the transformed, halfway space between the two longitudinal images, and applied an upper threshold of 1.2 to the resulting, halfway-transformed fractional anisotropy (FA) maps to remove noisy voxels. These thresholded FA maps were subsequently used in the tract-based spatial statistics (TBSS; Smith et al., 2006) processing approach for WM skeletonization.

1.2.4 Pre-processing – WMH/CSF Segmentation. The pipeline used FMRIB's Automated Segmentation Tool (FAST; Zhang et al., 2001) to segment the eroded, skull-stripped b_0 images into six separate maps based on voxel intensity (SI Figure 1). These included two images representing primarily WM in addition to images whose intensity reflected cerebrospinal fluid (CSF), areas of WMH, or image noise at interfaces between CSF and other tissue types, grey matter, and hypointense voxels reflecting noise at or near iron-containing subcortical nuclei (e.g., the basal ganglia). Fslmaths summed and binarized the two WM maps into a mask. Last, we refit the diffusion tensor in native space using the same procedures as above, but using the binarized WM mask. Next, fslmaths averaged the second and third eigenvalues created by dtifit for the WMH/CSF-masked data into radial diffusivity (RD) images for both occasions, and

renamed the first eigenvalue map as the axial diffusivity (AD) image.

1.2.5 TBSS-skeletonization. We used the TBSS processing framework to create a group-wise WM skeleton in standard space, and then to nonlinearly deproject the skeleton and WM atlases back to halfway space. We ran standard TBSS processing on the FA images generated by refitting the tensor in halfway transformed space. The FMRIB58_FA standard space image was used as the target for non-linear registration, and the data were nonlinearly aligned into 1 mm × 1 mm × 1mm MNI 152 space. The TBSS pipeline then generated the mean, standard space FA image from both waves of the sample, and the corresponding WM skeleton. We used a threshold of 0.3 in the final step of TBSS processing (i.e., tbss_4_prestats), to reduce areas with poor reliability in the mean WM skeleton. Last, the tbss_deproject routine was applied to the mean WM skeleton mask, and to the JHU-ICBM white matter atlas labels at 1mm, and the JHU-ICBM white matter tractography atlas (Mori et al., 2005; Wakana et al., 2007); the '2' and '–n' flags were used to nonlinearly warp the skeleton mask and atlases back to the space they were in at the first step of TBSS processing – here, the halfway space between measurements, while maintaining the integer values of the atlas regions. Furthermore, by using tbss_deproject on the atlases, the atlas values were only deprojected along the WM skeleton.

1.2.6 De-projection. We used FSL's FLIRT process to transform the atlas-derived, skeletonized regions of interest (ROIs) and the mean WM skeleton mask from individual halfway space to the original, native space from acquisition. Using fslmaths, we applied a lower threshold of 0.20 and an upper threshold of 1.001 to the native FA images from both occasions. This was necessary as FSL's tensor fitting method permits estimation of FA values over 1.0, primarily due to noise or motion. The thresholded FA image was then binarized and used to mask the deprojected skeleton mask in native space, reducing noise from each individually deprojected skeleton from which we sampled values. We used the FSL 'applywarp' function to deproject the Harvard-Oxford subcortical atlas (Desikan et al., 2006) to native space, using the subject-specific inverse warp matrices generated by TBSS and the halfway transformation

matrices created earlier; whereas tbss_deproject restricts deprojected values to the WM skeleton, this deprojection method does not.

1.2.7 Mask Creation. For each subject, we extracted separate masks from the atlases deprojected to native space. For redundancy, we used the individual native space, FA-thresholded skeletons as a secondary mask on the ROI masks during extraction. This eliminated any noisy voxels from individual ROIs not apparent on the group WM skeleton. When possible, we extracted separate masks for left and right and combined hemispheres.

In addition to mask extraction from the deprojected, skeletonized atlases, some additional manipulation of the masks was necessary. The atlases contained only individual masks for non-lateralized structures on the midline, such as corpus callosum (CC). Therefore, we used fslmaths to extract separate hemispheric masks from the Harvard-Oxford atlas transformed to individual subject space. We used these to mask left and right sides of the CC including body, genu, and splenium, to create separate lateralized masks. In addition, we sought to eliminate overlap between ROIs. The JHU-tractography atlas masks for forceps minor and forceps major were spatially redundant with the masks for CC genu and CC splenium, respectively. In order to model the variance from those regions separately, we used fslmaths to create separate, non-overlapping masks by subtracting the masks for genu and splenium from those for forceps minor and forceps major, respectively.

In addition, visual inspection of the uncinate fasciculus (UF) masks revealed substantial overlap between the two atlases in this rather small ROI. Therefore, we used fslmaths to sum the separate masks from the two atlases into a new UF mask with increased coverage.

1.2.8 Data Sampling. Based on visual inspection of the resultant masks, we chose 13 WM atlas-derived ROIs for sampling and analysis. The WM masks included the following tracts and regions previously associated with cognitive abilities in studies of aging: CC genu, CC splenium, CC body, dorsal cingulum bundle (CBd), ventral cingulum bundle (CBv), superior longitudinal fasciculus (SLF), and two divisions of the internal capsule, the anterior limb (ALIC),

and posterior limb (PLIC) – all taken from the ICBM-DTI-81 white matter labels atlas. In addition we included ROI masks taken from the JHU WM tractography atlas, including the inferior longitudinal fasciculus (ILF), inferior frontal-occipital fasciculus (IFOF), forceps major and forceps minor, as well as the UF mask created from both JHU atlases. Care was taken to ensure masks did not overlap. We chose to exclude masks that demonstrated visually apparent inconsistencies in coverage across the sample. These included the masks for superior frontal-occipital fasciculus, corticospinal tract, and corona radiata.

Data were sampled from the native space data masked for WMH/CSF for each of the atlas-derived masks. The fslstats function sampled and output the mean values and standard deviations for non-zero voxels for masks from left, right, and combined hemispheres for the three DTI indices AD, RD, and FA. We inspected standard deviations from the sampled FA values to help ensure no subject had excessive noise in a given ROI (e.g., standard deviations for all FA data were < .15). In addition, we evaluated the standard error of the mean values for each region sampled for each participant. Standard error values were low (FA SE range = 0.006 to 0.08), and varied according to the size of the region sampled.

1.3 Data Conditioning

One concern was cases included sufficient number of voxels for meaningful signal. In select cases, masking resulted in a relative dearth of coverage for six ROIs. Cases with voxels numbering ≥ 3 standard deviations below the mean were excluded from analysis. Thus, one case was removed from analysis of CC splenium, and CBd, two cases were excluded from models of CBv and ILF, and three cases were removed from analyses of UF.

1.4 Data Analysis

The paramount concern with the latent chance score modeling approach is the need to establish metric invariance (Meredith, 1964), a statistical precept necessary for longitudinal modeling of LVs in which the relationship between manifest and LVs must remain constant between measurement occasions. Meeting this assumption is necessary for confident

interpretation of the changes at the latent level. Metric invariance can be tested in Mplus by simply constraining the factor loadings for the two occasions to be equivalent. If there is metric invariance over time, the model provides a good fit for the data. However, if model fit borders on acceptable, there are small steps that can be taken in an attempt to modify the models. These include freeing factor covariances, factor means, or residual variances, fixing factor variances to 1, and fixing baseline factor means, individual paired indicator intercepts, and auto-correlated residuals between the baseline and change score factors to zero. In cases in which it does not reduce model fit, degrees of freedom can also be gained by imposing additional equality constraints on observed indicator intercepts and variance/residual variance, as well as on auto-correlated residuals for a given indicator between measurement occasions. If model fit is still unacceptable and can only be improved by removal of the equality constraints on the Time 1 and Time 2 factor loadings, then there is significant measurement variance over time, and change in that cognitive domain cannot be modeled.

2. Results

2.1 Exclusion of HDL cholesterol from metabolic risk factor models

Initially, using data for the normotensive-only sample and thus avoiding confounding effects of medication, we fitted a CFA model to the baseline metabolic risk indicators: log-transformed triglyceride level, HDL cholesterol level, systolic blood pressure, waist-to-hip ratio, and fasting blood glucose level. HDL was multiplied by -1 to align its scaling with the other indicators. The fit was acceptable only if bidirectional paths were specified between HDL and both triglycerides and waist-to-hip ratio ($\chi^2[2] = 0.41$, $p = .813$, CFI/TLI = 1.00/1.10, RMSEA = 0.00, SRMR = 0.01). Moreover, the univariate LCSM for metabolic syndrome using the same indicators as those in the CFA and their longitudinal counterparts also fit adequately ($\chi^2[47] = 56.77$, $p = .156$, CFI/TLI = 0.99/0.98, RMSEA = 0.05, SRMR = 0.08), even without specifying the correlational paths for HDL. Furthermore, there was significant variance in the latent change score (estimate/S.E. = 2.13, $p = .034$). However, when we subsequently incorporated the

metabolic syndrome risk (Met) factor LCSM in multivariate models to evaluate the effects of change in vascular risk on DTI indices, the models did not fit well. Model fit improved after specifying correlational paths between HDL and triglycerides, which violated basic requirements of the LCSM framework. Thus, we specified an alternative LCSM for vascular risk without HDL in the model. The new LCSM with metabolic risk score without HDL was a good fit for the data (χ^2[31] = 21.90, p = .886, CFI/TLI = 1.00/,1.03 RMSEA = 0.00, SRMR = 0.06), but the estimated variance in the change score was no longer significant (estimate/S.E. = 1.79, p = .074).

References

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006;31(3):968-980.

Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal. 2001;5(2):143-156.

Meredith W. Notes on factorial invariance. Psychometrika. 1964;29(2):177-185.

Mori S, Wakana S, Nagae-Poetscher LM, van Zijl PC. MRI Atlas of Human White Matter. Amsterdam: Elsevier B.V.; 2005.

Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage. 2006;31(4):1487-1505.

Tabachnick, B. G., and Fidell, L. S. (2013). *Using Multivariate Statistics , 6th ed.* Boston : Pearson.

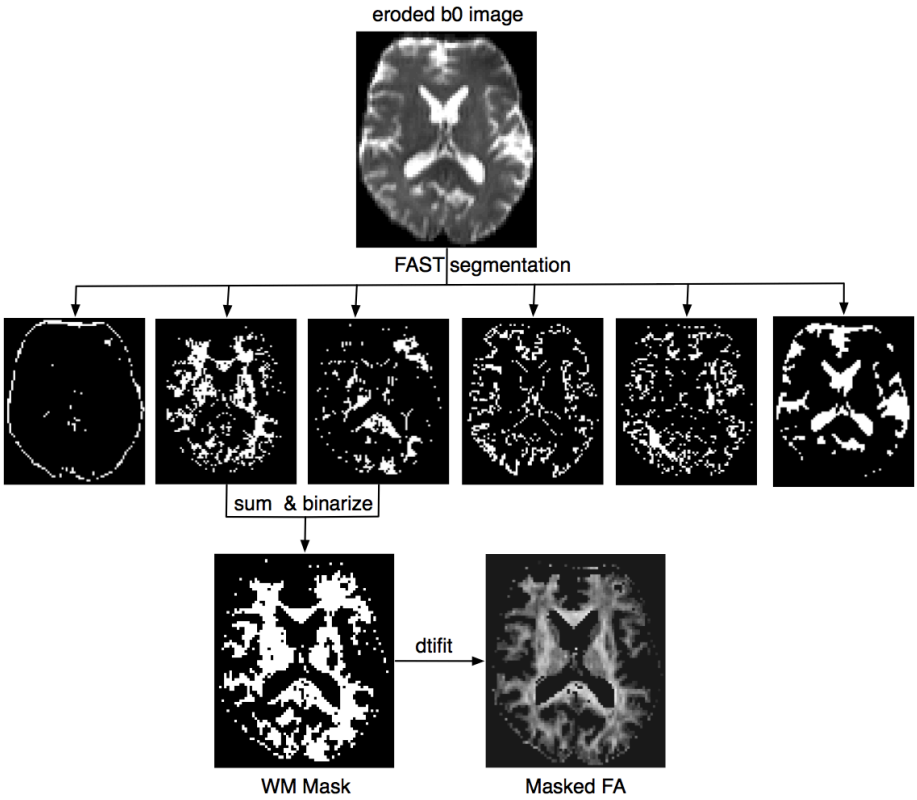
Wakana S, Caprihan A, Panzenboeck MM, Fallon JH, Perry M, Gollub RL, et al. Reproducibility of quantitative tractography methods applied to cerebral white matter. Neuroimage. 2007;36(3):630-644.

Zhang YY, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging. 2001;20(1):45-57.

SI Table 1. Test-retest correlations between Time 1 and Time 2 for unstandardized values and latent factors

	CBd	CBv	IFOF	ILF	SLF	UF	CC Genu	CC Body	CC Splenium	Forceps Major	Forceps Minor	ALIC	PLIC
FA													
Unstd. Left	0.441	0.679	0.765	0.652	0.795	0.631	0.822	0.538	0.594	0.628	0.813	0.774	0.467
Unstd. Right	0.202	0.414	0.832	0.581	0.889	0.706	0.843	0.534	0.648	0.760	0.810	0.658	0.642
Latent factor	0.620	0.999	0.918	0.978	0.962	0.884	0.998	0.666	0.958	0.986	0.863	0.947	0.894
AD													
Unstd. Left	0.445	0.315	0.549	0.638	0.777	0.681	0.473	0.147	0.467	0.468	0.700	0.778	0.674
Unstd. Right	0.153	0.346	0.640	0.385	0.820	0.510	0.418	0.023	0.334	0.542	0.695	0.780	0.619
Latent factor	0.711	0.647	0.941	0.967	0.926	0.971	0.975	0.369	0.968	0.823	0.887	0.950	0.996
RD													
Unstd. Left	0.560	0.471	0.846	0.726	0.840	0.594	0.820	0.787	0.675	0.727	0.843	0.852	0.426
Unstd. Right	0.367	0.365	0.868	0.778	0.897	0.535	0.867	0.788	0.792	0.762	0.838	0.786	0.584
Latent factor	0.888	0.986	0.969	0.999	0.974	0.862	0.999	0.841	0.949	0.997	0.913	0.949	0.773

Notes: Unstd. Left and right are raw mean values sampled from ROIs before standardizing data to z-scores. Latent factor scores were saved from univariate latent change score model output.



SI Figure 1. Diagram depicting the procedure used for segmenting out cerebral spinal fluid and white matter hyperintensities. Using FSLs FAST tool for segmenting the b_0 image into 6 images based on intensity, the two maps corresponding to white matter (WM) were summed, binarized, and the tensor refit in the WM-only mask for subsequent sampling.