# REGISTRATION AND SEGMENTATION OF BRAIN MR IMAGES FROM ELDERLY INDIVIDUALS

by

**Minjie Wu**

B.S., Wuhan University, 1998

M.S., Wuhan University, 2001

M.S., University of Pittsburgh, 2008

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Minjie Wu

It was defended on

November 17, 2009

and approved by

Howard J. Aizenstein, M.D., Ph.D., Associate Professor, Department of Psychiatry and
Bioengineering, University of Pittsburgh

George Stetten, M.D., Ph.D., Professor, Department of Bioengineering and Radiology,
University of Pittsburgh; Robotics Institute, Carnegie Mellon University

Theodore Huppert, Ph.D., Assistant Professor, Department of Radiology and Bioengineering,
University of Pittsburgh

Carmen Andreescu, M.D., Assistant Professor, Department of Psychiatry, University of
Pittsburgh

Dissertation Director: Howard J. Aizenstein, M.D., Ph.D., Associate Professor, Department
of Psychiatry and Bioengineering, University of Pittsburgh

**THE REGISTRATION AND SEGMENTATION OF BRAIN MR IMAGES FROM**

**ELDLY INDIVIDUALS**

Minjie Wu, Ph.D.

University of Pittsburgh, 2009

Quantitative analysis of the MRI structural and functional images is a fundamental component in the assessment of brain anatomical abnormalities, in mapping functional activation onto human anatomy, in longitudinal evaluation of disease progression, and in computer-assisted neurosurgery or surgical planning. Image registration and segmentation is central in analyzing structural and functional MR brain images.

However, due to increased variability in brain morphology and age-related atrophy, traditional methods for image registration and segmentation are not suitable for analyzing MR brain images from elderly individuals. The overall goal of this dissertation is to develop algorithms to improve the registration and segmentation accuracy in the geriatric population. The specific aims of this work includes 1) to implement a fully deformable registration pipeline to allow a higher degree of spatial deformation and produce more accurate deformation field, 2) to propose and validate an optimum template selection method for atlas-based segmentation, 3) to propose and validate a multi-template strategy for image normalization, which characterizes brain structural variations in the elderly, 4) to develop an automated segmentation and localization method to access white matter integrity (WMH) in the elderly population, and finally 5) to study the default-mode network (DMN) connectivity and white matter hyperintensity in late-life depression (LLD) with the developed registration and segmentation methods.

Through a series of experiments, we have shown that the deformable registration pipeline and the template selection strategies lead to improved accuracy in the brain MR image registration and segmentation, and the automated WMH segmentation and localization method provides more specific and more accurate information about volume and spatial distribution of WMH than traditional visual grading methods. Using the developed methods, our clinical study provides evidence for altered DMN connectivity in LLD. The correlation between WMH volume and DMN connectivity emphasizes the role of vascular changes in LLD's etiopathogenesis.

**Keywords:** image registration, image segmentation, template selection, resting state connectivity, late-life depression.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF NOTATION (ACRONYMS)

# PREFACE

The accomplishment of this Dissertation would not have been possible without the support of many individuals. First and foremost, I am especially grateful to my advisor and mentor, Dr. Howard J. Aizenstein for the guidance, encouragement, tremendous support and opportunities he provided throughout my graduate studies at the University of Pittsburgh.

I would like to express my greatest appreciation and gratitude to my doctoral committee members, Drs. George Stetten, Theodore Huppert, and Carmen Andreescu for their time, insightful comments, and constructive suggestions on my work.

I would like to thank all my colleagues at the Geriatric Psychiatry Neuroimaging Lab for their help, support and most importantly their friendship. It has been an enjoyable experience working in a group with such professionalism and creativity.

Finally, I wish to thank my husband, my parents, my brother and sister for their unconditional love and encouragement throughout this journey.

# 1.0     INTRODUCTION

Magnetic resonance imaging (MRI) plays an increasingly important role in the studies of the brain in both modern neuroscience and radiological practice. This non-invasive imaging technique is flexible and sensitive to a broad range of tissue properties, which allow for in vivo studies of the anatomy, the function, and the metabolism of the human brain. Brain anatomy is studied via quantitative analysis of the MRI structural images. Studying anatomical brain structures is important for assessing the brain's anatomical abnormalities, mapping functional activation onto human anatomy, longitudinally evaluating disease progression, and planning computer-assisted neurosurgery [1]. Brain function is studied via functional Magnetic Resonance Imaging (fMRI). One use of this technique is to help measure the localized temporal haemodynamic response triggered by neural activity in the brain [2]

Image registration and segmentation is central in analyzing structural and functional MR brain images. However, due to increased variability in brain morphology in the elderly, traditional methods for image registration and segmentation of MR brain images are sometimes not suitable for analyzing MR brain images of elderly individuals. The goal of this dissertation is to develop algorithms to improve the registration and segmentation accuracy in the MR images of the geriatric population.

## 1.1  STANDARD APPROACHES IN BRAIN MRI

Many approaches have been developed for brain MRI research in the elderly. These include utilizing a variety of MR acquisition contrasts, as well as different image registration and segmentation methods for image data analyses.

### 1.1.1  MR Contrast

In neuroscience and radiological applications, the MR signal most frequently relies on the relaxation property of excited hydrogen nuclei in the water molecules and lipids. The spatial encoding of the MR signal is accomplished by superimposing gradient fields at three directions, x, y and z. The contrast of an MR image is dependent on the MR imaging protocol. Applying radio-frequency (RF) pulses and gradient pulses at carefully chosen timings may produce images with different intensity contrasts, which highlight different tissues in the brain.

Using the intrinsic magnetic properties, three types of image contrasts exist for MR imaging of tissue: T1-weighted, T2-weighted and proton density. MRI can also detect the intrinsic features of tissue microstructure and microdynamics using diffusion-tensor MRI (DT-MRI). A montage of a typical MR image at an axial slice, with various MR image types of acquisition including T1, T2, proton density weighted images, is shown in Figure 1.1. Echo-planar-based diffusion weighted image, anisotropy and tensor orientation maps from diffusion-tensor imaging are also included in Figure 1.1.
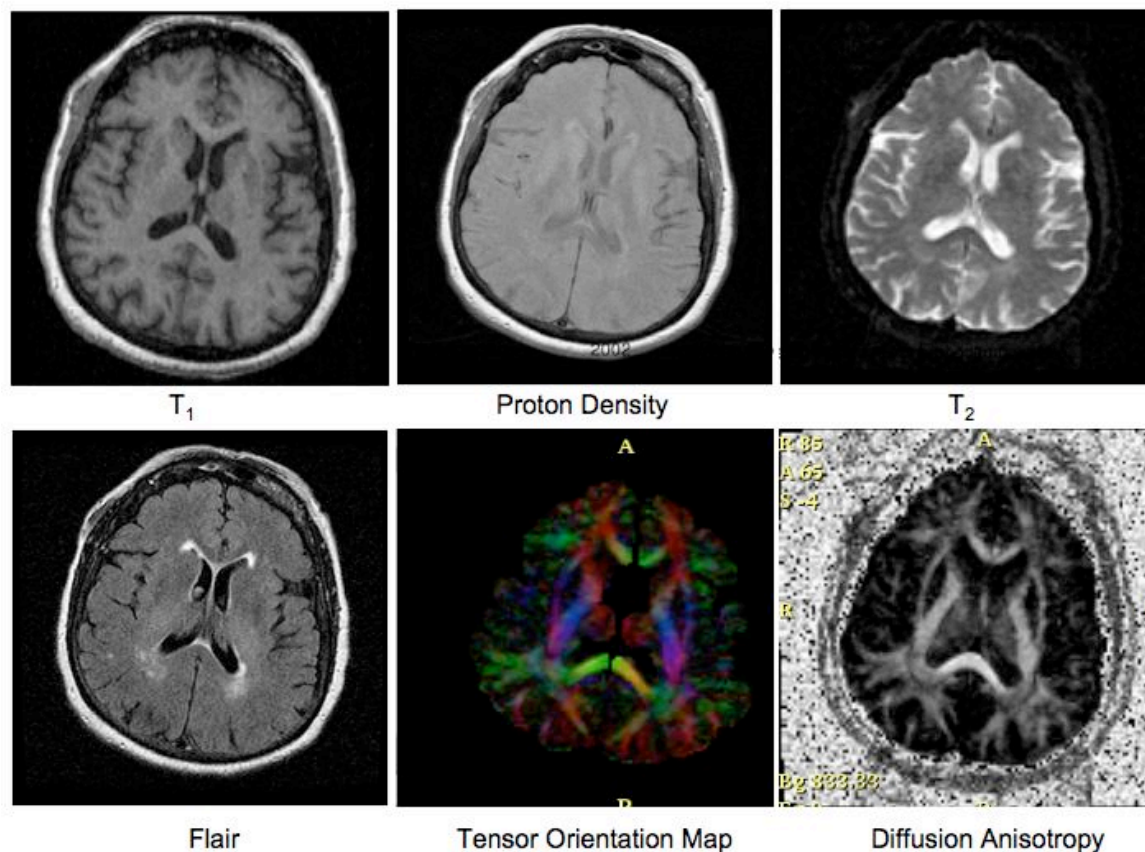
**Figure 1.1. A variety of types of MR images can be acquired for brain imaging.**

## 1.1.2 Image Registration

Image registration involves aligning one image to another image acquired for different objects or the same object but under different conditions. The process requires finding the optimal spatial transformation that maps corresponding structures in one image to the other. This is both important and ubiquitous in many aspects of medical imaging; particularly in the analysis of MR brain imaging. For example, image registration can be used to transform individual brain images into the standard stereotactic image space, so as to facilitate the comparison across subjects in fMRI. Image registration can also be used in atlas-based image segmentation to automatically

label the anatomy of new brain images. Depending on what types of features are used in transforming the template image into the subject image space, a given image registration method can be classified as either model-based or intensity-driven.

**1.1.2.1 Model-based registration**

Model-based registration approaches involve matching the homologous landmarks or features [3]. These landmarks or features may be obtained manually or semi-automatically. The advantage of the model-based approaches is that images with different modalities can be registered using the extracted landmarks; but manually determining these landmarks could be very labor intensive (to the point of infeasibility) and often requires extensive training.

**1.1.2.2 Intensity-based registration**

Intensity-based approaches obtain the spatial transform by maximizing a similarity criterion between the two images [4-6]. The spatial transformation describes the spatial relationship between the target image and subject image; the metric (similarity criterion or cost function) quantitatively describes how well the target image and subject image are registered; the optimization method decides the procedure used to search the space of transformation parameters to find an optimal transformation as a function of the metric.

*Spatial transformation models*

Several spatial transformation models have been proposed to estimate the deformation field from the source image to the target image, including affine linear model [7], nonlinear polynomial model [7], linear combination of basis functions [6], elastic registration technique [8-10], demons registration model [11] and so on.

The 12-parameter (9 independent parameters) affine linear model requires that parallel lines remain parallel after transformation, which allows rotation, translation, rescaling and shearing along the three axes.

In the polynomial model, the transformed coordinates $(y_{i1}, y_{i2}, y_{i3})$ in the target image space are computed as a polynomial function of respective location $(x_{i1}, x_{i2}, x_{i3})$ in the source image.

To estimate the deformation field, the nonlinear spatial transformation model is expanded as a linear combination of some smooth basis functions $\beta(x_1, x_2, x_3)$, such as 3D discrete cosine transform (DCT) functions [6].

Elastic registration techniques model the deformation of the source image into a target image as the physical deformation of an elastic body under external forces $f(x, y, z)$. The physical deformation involves internal and external forces. The internal force tends to resist any change in shape. The deformation of the elastic body stops when the internal force and the external force form an equilibrium state.

Demons registration models the image registration problem as a diffusion process, in which the contour (surface) of the target image is considered as a membrane and the deformed image is considered as a deformable grid.

***Similarity metrics***

A number of similarity metrics have also been explored for the intensity-based MR brain warping. These include least squared differences in voxel intensities [4], ratio image uniformity (RIU) [4], normalized cross-correlation [9], and mutual information [12].

With least squared difference metric, the difference between the source image and the target image is computed at each voxel, and the square of this difference is averaged across voxels to generate this cost function. The metric assumes that both images are of the same modality and of similar intensity distribution. Intensity histogram matching is needed to normalize the intensity scales of the source image and the target image when global intensity differences exist between two images.

For the ratio image uniformity metric, a ratio image is created by dividing the source image by the target image at the voxel level. Then, the standard deviation of the ratio image normalized by the mean ratio is calculated as the measure for the ratio image uniformity. When there is global intensity scaling between the target image and the source image, the registration performance with RIU metric improves.

The mutual information metric is measured from the joint probability distribution of the images' intensities, which evaluates the statistical dependency between the intensity patterns of the source image and the target image. The main advantage of this metric in the brain warping is that it is suitable for both multi-modality and same-modality image registration.

Image registration is used in atlas-based segmentation to automatically segment anatomical structures on a given subject's brain image, in which the template image is transformed into the subject space (template->subject), carrying the labeling atlas along from the template space into the subject image space. Image registration is also used in image normalization to warp the individual brain image into a common standard space (subject->template), which facilitates voxel-wise comparison between groups in fMRI.

### 1.1.3   Image Segmentation

Image segmentation plays an important role in many neuroimaging applications including neurosurgical planning, diagnosis and assessment of various pathologic conditions, and the definition of anatomical regions of interest for functional imaging studies. Image segmentation involves assigning a label to every voxel in an image, such that all the voxels in a segment with the same label share certain characteristics. Automated and semi-automated segmentation methods that have been developed for image segmentation include clustering methods, histogram-based methods, neural networks segmentation, and atlas-based segmentation.

**1.1.3.1 Atlas-based segmentation**

Accurate structural or regional segmentation and classification is critical for many MR brain image analyses such as volumetric comparison or inter-subject functional activation comparisons. Traditional manual region segmentation is not only labor intensive and time-consuming (to the point of infeasibility with a large dataset and multiple regions), but also introduces human subjectivity and often requires extensive training [13]. To overcome these drawbacks, atlas-based segmentation was introduced to automatically label the anatomical structures for individual brain images [9, 14]. In atlas-based segmentation, the standard labeled atlas brain image is registered to the individual brain image using different spatial deformation models, and then the labeling information in the atlas is carried into the subject space by warping the atlas with the obtained spatial transformation.  In this method, the region segmentation or classification task is viewed as a registration procedure, and the key problem becomes finding the optimal spatial transformation between the template brain image and the individual brain image.

As show in **Figure 1.2**, the standard Montreal Neurological Institute (MNI) template brain Colin27 [15], which carries high anatomical details and has a high spatial resolution (1mm x 1mm x 1mm voxel size), is used as the template. In the figure, an axial image slice of the Automated Anatomical Labeling digital atlas (AAL) [16] on MNI Colin27 and one elder subject are also shown.
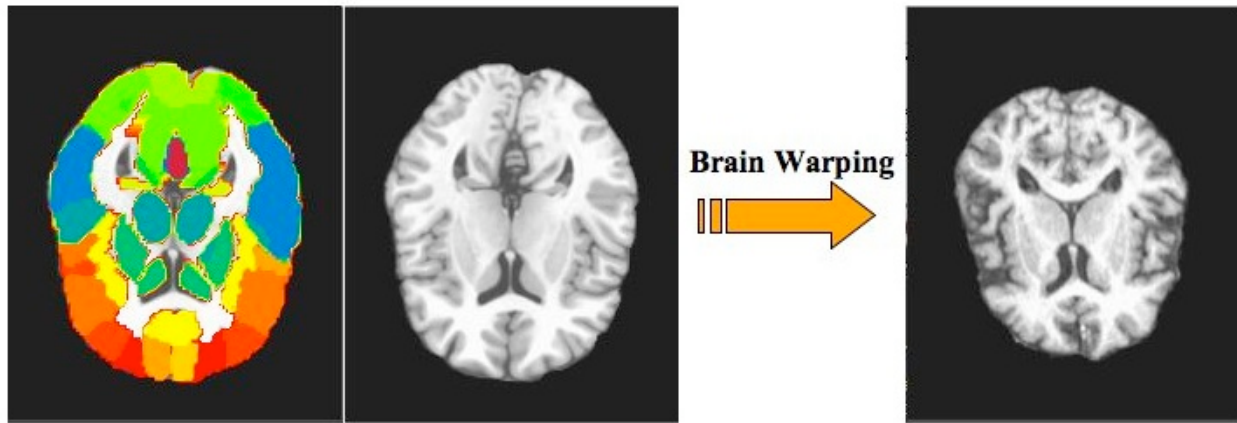


**Figure 1.2. Axial slice of structural MR brain images of MNI template colin27, AAL atlas and one elderly subject.**

## 1.1.4    Image Normalization

In image normalization, functional and structural data from different individuals are warped into a stereotaxic coordinate space (e.g. MNI Colin27) using different registration methods (subject->template image registration). Image normalization is a prerequisite in most neuroimaging research for voxel-wise functional or structural image analyses.

For example, when studying brain function within a population or comparing brain function between different groups, the motion-corrected low-resolution functional images are first registered into an individual structural image using rigid body or linear affine model, then

they are warped into MNI colin27 using a fully deformable model, and finally a voxel-by-voxel t-test is performed on the warped functional images (in standard image space) to locate functional activation, as shown in Figure 1.3. If the same stereotaxic coordinate space (e.g. MNI Colin27) is used for different studies, then the anatomical location of functional activation can be reported as the standardized coordinates, and with this quantitative spatial reference system results (e.g. functional activation maps) from different studies or groups can easily be compared and correlated.
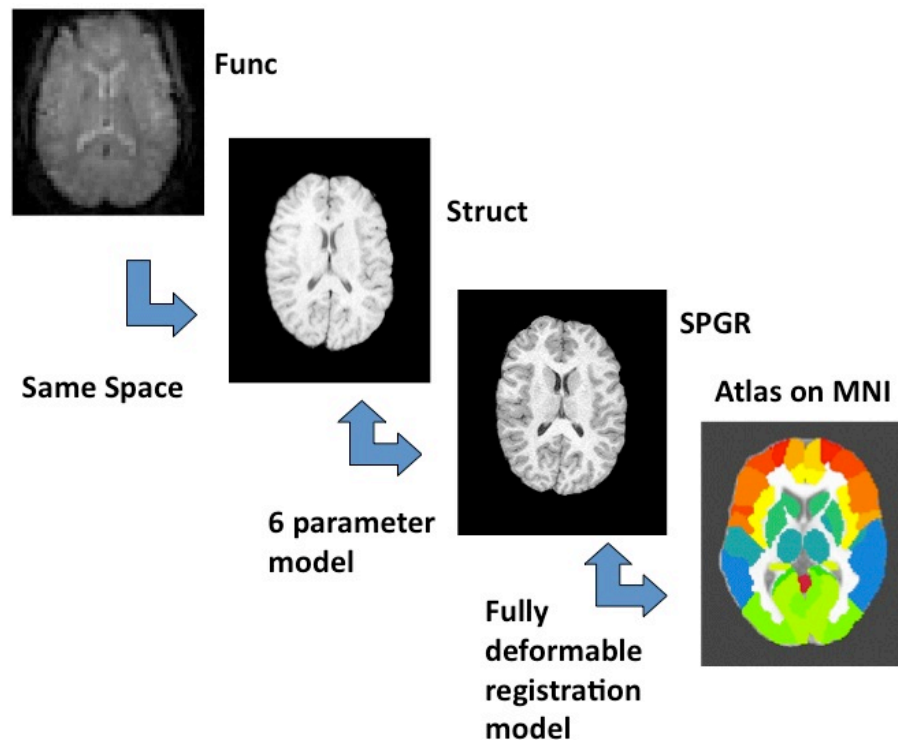


**Figure 1.3. Image normalization to convert functional MR images of individuals into a common space MNI Colin27. Func: low-resolution functional brain image, Struct: low-resolution structure brain image acquired at the same location as functional image, SPGR: Spoiled Gradient Recall Acquisition, high-resolution T1-weighted brain image, and MNI atlas: Montreal Neurological Institute standard template Colin27.**

## 1.2    BRAIN MORPHOLOGY IN THE ELDERLY

### 1.2.1    Variations in Normal Brains

Individual brains in a population do not conform to a single anatomic topology. Normal variations in the brain structures include difference in sulcal patterns, gyral folding patterns, morphological shape, volume, asymmetries between the right and left hemispheres, etc. For example, variability in the gyral folding pattern is present in the anterior cingulate cortex (ACC) structure; it has been estimated that approximately 30-60% of the population has a paracingulate sulcus (PCS)  [17], a normal variant of the ACC in which there is an additional gyral fold (see Figure 1.4).
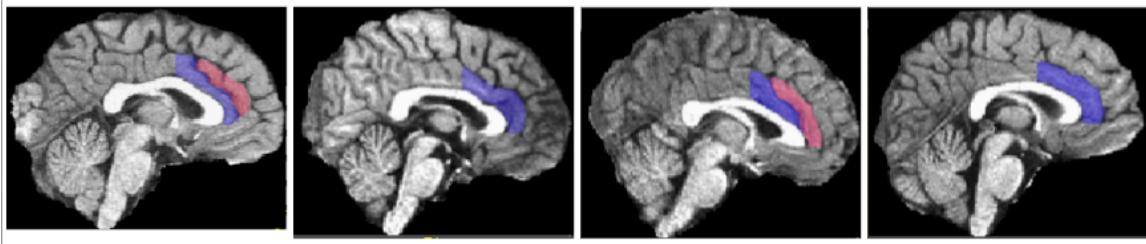


**Figure 1.4. The variations in anterior cingulate cortex (ACC) are present in the brains of 4 healthy subjects. Two subjects have an additional gyral fold (red) paracingulate sulcus (PCS).**

### 1.2.2    Variations in Elderly Brains

Due to age-related atrophy, elderly brains have smaller brain volume, larger sulci and larger ventricles. The presence of these significant differences is shown in Figure 1.5, in which an elderly brain is compared to the standard template (e.g. MNI Colin27).
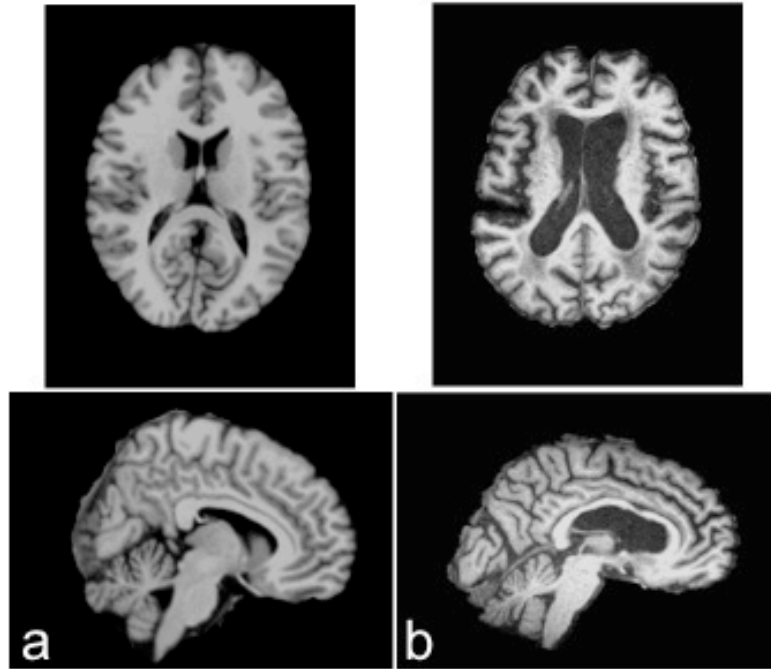
**Figure 1.5. The comparison of brain images from (a) young subject versus (MNI Colin27) (b) elderly subjects.**

The magnitude of brain shrinkage during aging varies among brain regions and structures. Longitudinal and cross-sectional studies have shown more age-related pronounced tissue loss in frontal and parietal than temporal and occipital lobes [18]. However, a different brain aging pattern has been reported by Raz et al., in which severe and moderate age-related atrophies are observed in prefrontal and temporal cortices while minor differences are observed in parietal and occipital cortices [19]. The local patterns of gray matter and white matter loss during aging also differ. Age-related gray matter change has been found to be more prominent in cingulate, insular, orbital and inferior frontal cortex than medial temporal regions [18]. However, white matter volume loss has been observed to be more widespread throughout the elderly brains [18]. Figure 1.6 shows a typical T2-weighted attenuated inversion recovery (FLAIR) image of the same elderly subject with severe white matter hypertensies (WMH, also known as leuokoaraiosis).
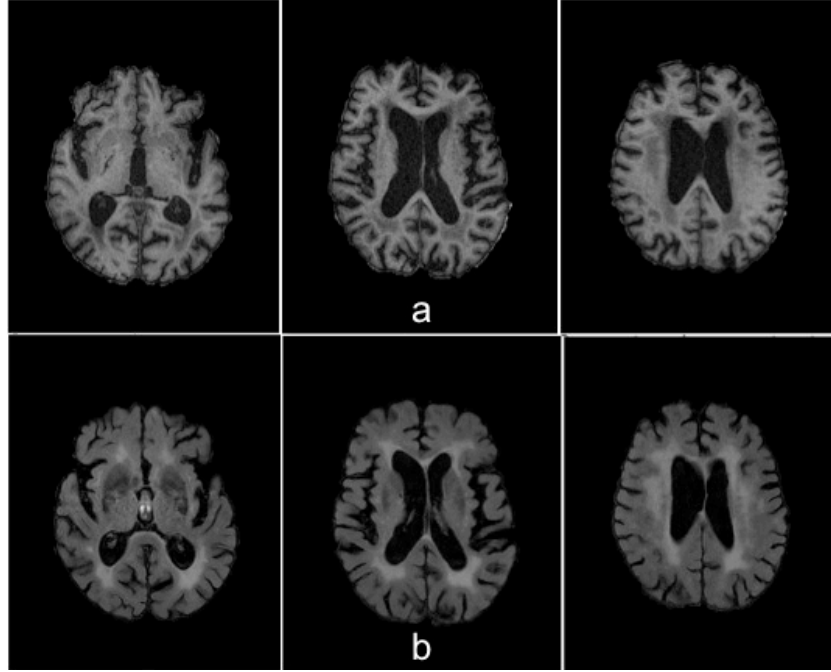
**Figure 1.6. The T1-weighted MR brain image (a) and T2-weighted FLAIR image (b) with severe white matter hyperintensities due to demyelination in an elderly subject.**

The rate of brain shrinkage with age also varies throughout the brain. Previous research shows linear correlations between age and loss of tissue in the cortex, amygdala, thalamus, accumbens, and caudate; it also shows curvilinear relationships between age and shrinkage of the cerebral white matter, hippocampus, brainstem, cerebellar white, gray matter, and so on [20].

## 1.3     MOTIVATION AND GOAL

In addition to the variations in normal brain anatomical structures, the heterogeneity in age-related brain atrophy introduces additional variability in the brain morphology of elderly individuals. This presents significant difficulties in the image registration and segmentation of the MR brain images from elderly population. In this dissertation, possible solutions are

proposed to improve the registration and segmentation accuracy of MR brain images of the elderly.

### 1.3.1 Comparison of Registration Methods

The standard nonlinear registration algorithms such as Automated Image Registration (AIR)[21] and Statistical Parametric Mapping (SPM)[6] either allow linear or nonlinear smooth deformation fields or are restricted to small deformations, which may not be sufficient to correct the highly nonlinear difference between the elderly subject and the template MNI Colin27 in image alignment. The first goal of this dissertation is to implement a fully deformable registration pipeline to allow a higher degree of spatial deformation, and to quantitatively compare the performance of the deformable model to the popular linear and nonlinear registration methods (AIR and SPM) in atlas-based segmentation.

### 1.3.2 Template Selection Strategies

Using an unrepresentative single brain template in atlas-based segmentation or image normalization [e.g. the Talairach atlas [22], the MNI Colin27 [15]] introduces a bias in the image registration by favoring subjects similar to the template against subjects substantially different from the template. A single-brain template, MNI Colin27 for example, is not a good representative of geriatric populations given the increased variability in brain morphology in the elderly. Different template selection strategies have been proposed to improve the accuracy in atlas-based segmentation and image normalization.

**1.3.2.1 Template Selection in Atlas-based Segmentation**

A second goal of this dissertation is to improve the segmentation accuracy with an optimum template selection strategy. Instead of choosing a fixed brain atlas, we use a family of brain templates for the atlas-based segmentation. For each subject and each region the template selection method automatically selects the 'best' template with the highest local region of interest (ROI) registration accuracy based on the metric of normalized mutual information. The segmentation performances of the optimum template selection method and the single template method are quantitatively evaluated and compared.

**1.3.2.2 Template Selection in Image Normalization**

Another goal of this dissertation is to propose multi-template strategies to improve the registration accuracy in the image normalization of elderly population. Instead of registering an individual brain image directly into the standard MNI space Colin27 during image normalization, this method chooses intermediate prototype templates as a bridge between the individual brain image and Colin27 (subject->intermediate brain images->Colin27). In order to identify a set of intermediate templates, we use a bottom-up clustering algorithm to choose a set of prototype templates that spans the space of brain morphometry in the elderly population [23-26]. The purpose of this strategy is to characterize the anatomical structural variations in brains of the elderly population while preserving brain structural details.

**1.3.3   Segmentation and Localization of White Matter Hyperintensities**

White matter hyperintensities (WMH), commonly seen in T2-weighted FLAIR MR images, reflect aged-related white matter damage. As a marker for white matter degeneration, WMHs are

associated with neuropsychiatric disorders, including vascular dementia [27], Alzheimer's disease [28], and late-onset late-life depression [29, 30]. Previous MRI studies of WMHs have primarily relied on the subjective and global (i.e., full-brain) ratings of WMH grade. Another goal of this dissertation is to implement and validate an automated method for quantifying and localizing WMHs. This automated segmentation method is then used to study the effect of WMH burden on the resting-state connectivity in late-life depression.

### 1.3.4 Clinical Application

Using the fully deformable registration and the automated WMH segmentation methods developed in the previous sections, we characterize the default-mode network (DMN) connectivity in late-life depression (LLD) and the correlation of default-mode network (DMN) activity changes with the White Matter Hyperintensity (WMH) burden. We hypothesize that LLD subjects will have altered DMN activity, which will correlate with the increased WMH burden, providing evidence for the vascular depression hypothesis.

## 1.4 DESCRIPTION OF CHAPTERS

Overall, the goal of this dissertation is to improve the accuracy of image registration and segmentation of elderly population. The first chapter gives an overview of standard approaches in brain MRI and challenges in analyzing MR brain images from elderly individuals due to age-related atrophy.

The second chapter focuses on developing a deformable pipeline to achieve more accurate estimation of the deformation fields (template Colin27->Subject). The performance of the deformable model is quantitatively compared to the popular packages in the neuroimaging field through a series of experiments.

The third chapter proposes an optimum template selection strategy to further improve the segmentation accuracy in atlas-based segmentation and compares the proposed template selection method to the standard single-template strategy.

The fourth chapter addresses the template selection problem in image normalization of elderly population. A multiple template strategy is proposed and evaluated under different spatial deformation models.

The fifth chapter proposes an automated method to segment and localize the white matter damages (white matter hyperintenties) in T2-weighted FLAIR images of the elderly population.

The sixth chapter focuses on the clinical applications of the previously developed methods, and studies the effect of white matter damage on default-mode network in late-life depression.

The final chapter discusses the limitations and contributions of this dissertation. Potential directions for further research are also explored.

# 2.0     COMPARISON OF REGISTRATION METHODS

Typical packages used for co-registration of neuroimages in functional image analyses are Automated Image Registration [AIR, 4] and Statistical Parametric Mapping [SPM, 6]. However both methods have limited-dimension deformation models; AIR uses a polynomial transformation model with limited coefficients, while SPM uses the linear combination of smooth basis functions. On the other hand, a fully deformable registration technique [31] which combines the piecewise linear registration for coarse alignment with the demons algorithm [11] for finer tuning, allows for a higher degree of deformation and a more accurate spatial deformation field. Moreover, the fully deformable technique is publicly available (www.itk.org), like AIR and SPM. Recent work by our group suggests that a similar fully deformable method [32] is more accurate than affine linear methods at hippocampus segmentation in Alzheimer's disease patients [33]. In this chapter, we quantitatively compare the performance of AIR, SPM and the fully deformable model in three aspects with control subjects: accuracy of automatic segmented region against the ground truth region (i.e., hand-drawn region), the sharpness of the average brain image, and reliability of functional signal for group analysis.

## 2.1    MATERIALS

### 2.1.1   Subjects and MR Imaging Parameters

Ten subjects (7 male; mean age 24.4, range 20-32 years old; right-handed) participated in this study. Scanning was done on a 1.5T GE CVi scanner with 3D SPoiled Gradient Recall Acquisition (SPGR, TE/TR=5/25; flip angle = 40; FOV= 24×18cm, slice thickness 1.5mm, 256×192 matrix). Functional scanning was performed using a one-shot spiral sequence (TR/TE = 2000/35; flip angle = 70; FOV = 24; slice thickness = 3.8mm; 64×64×26 matrix). Subjects performed 8 blocks of a learning task, in which the stimuli appeared in one of four boxes across the screen and the subjects were asked to respond as fast and accurately as possible to the location of the stimuli with a key press (using the index and middle fingers of the both hands). An error for this task is defined as any trial in which the subject pressed the incorrect key while responding to the location of a stimulus. The stimuli appeared once every 2 seconds for 40 seconds, followed by 20 seconds of fixation.

### 2.1.2   Manual Segmentation of Brain Regions

The standard MNI (Montreal Neurological Institute) brain colin27 [15] which carries high anatomical details and has a high spatial resolution (1mm x 1mm x 1mm voxel size), was used as the template. Two raters manually segmented the right hippocampus and right anterior cingulate cortex (ACC) on the template colin27 and on each subject. The segmented regions on the template were used as the atlas in atlas-based automatic segmentation [34] and the segmented regions on each subject were used as the ground-truth region mask.

For the segmentation procedure, the hippocampus was manually traced in the sagittal view. The first medial slice was defined as the slice that first showed the cerebral peduncle separated from the upper pons and the most lateral slice was defined as the last slice that still showed gray matter of the hippocampus. The posterior limit was set as the slice where an ovoid mass of gray matter started to appear inferiomedially to the trigone of the lateral ventricle. The alveus served as the anterior and superior limit of the head of the hippocampus.

For the segmentation of the ACC, tracings were made in serial coronal slices. The sagittal and axial views were used as a reference to outline the ACC .The posterior limit of the ACC was defined by a vertical line perpendicular to the anterior commissure-posterior commissure (AC-PC) plane and passing through the AC. The cingulate and callosol sulci constituted of the outer and inner boundaries of the ACC respectively. When a sulcus running parallel and superior to the cingulate sulcus was present, the paracingulate gyrus was included in the tracing.

Inter-rater reliability for the manual tracings of each one of the two regions of interest (ROI) was calculated using the intraclass correlation coefficient (ICC). For the inter-rater reliability, the calculated ICCs were: 0.89 (right hippocampus) and 0.97 (right ACC). To obtain intra-rater reliability a subset of 5 MR images was retraced by the same rater after 3-4 weeks (mean 22.2 ±3.4 days). The ICCs for intra-rater reliability were: 0.99 (right hippocampus) and 0.93 (right ACC).

## 2.2 METHODS

### 2.2.1 Registration Methods

This section evaluates the image registration accuracies of three methods: AIR, SPM, and the fully deformable model. The registration methods were used to co-register the 3D SPGR image of each subject and the template Colin27. Prior to the registration, the skull was stripped from both colin27 and the subject's 3D SPGR using the Brain Extraction Tool [BET, 35]. For AIR, the registration procedure that we used starts with a 12-parameter affine linear registration and then is followed by a second order 30-parameter nonlinear polynomial model. This procedure appeared to produce the best registration accuracy [34, 36]. For SPM, we used the standard SPM registration method, which begins with an affine linear registration similar to AIR, and then proceeds with the nonlinear registration using the spatial transformation model consisting of a linear combination of low-spatial-frequency discrete cosine transform functions. Both methods use limited parameters to describe the spatial deformation field; hence they only allow a certain degree of spatial transformation, which may lead to inaccurate alignment between the individual brain image and the template due to local anatomical variability or pathologic brain changes. The fully deformable model in this study is similar to that used by Chen [32]. This model was implemented using the registration library from Insight Segmentation and Registration Toolkit (ITK). This method first uses with a grid-based piecewise linear registration, followed by the demons registration algorithm as a fine-tuning procedure for a voxel-level spatial deformation. The fully deformable registration allows for more spatial deformation, which seems to give it a particular advantage over the other two packages when the brains are significantly different from the atlas like the brains of the elderly population.

**2.2.2   Evaluation Methods**

In this section, we describe the three experiments that were performed to evaluate the relative accuracies of AIR, SPM, and the fully deformable model.  These experiments test the accuracies of registration using 1. atlas-based segmentation of the hippocampus and anterior cingulate cortex, 2. the smoothness of a mean image generated using the three registration approaches, and 3. the effect-size of the blood-oxygen-level dependent (BOLD) fMRI signal co-localized across subjects with these registrations.

**2.2.2.1 Atlas-based Segmentation (Experiment 1)**

In this experiment, the automatic segmented regions were compared to the manually labeled ground-truth region masks for all of the 10 subjects. The right hippocampus and right anterior cingulate cortex (ACC) of each subject were estimated through atlas-based segmentation (template->subject registration) with AIR, SPM, and the fully deformable model respectively. The overlap ratio—the ratio of structure voxels (*e.g.*, hippocampus voxels) that the estimated and ground-truth regions have in common to the structure voxels of the overall region (manually-drawn region + auto-segmented region)–is computed for each subject to quantify how well the auto-segmented anatomical structures overlap with the hand-drawn ground-truth region mask.

**2.2.2.2  Image Normalization (Experiment 2)**

In Experiment 2, ten individual brain images were warped into the template Colin27 using each registration approach, and then a mean brain image was created from the resulting warped images. Visual inspection and quantitative smoothness measurement of the mean brain image were used to evaluate the performance of the corresponding registration method across subjects

in image normalization. Misalignment or error from the registration leads to a blurred mean brain image. Differences due to variability in individual anatomy also impact registration accuracy. The brain images of the ten subjects are presented in Figure 2.1 to show the variability of individual brain anatomy.
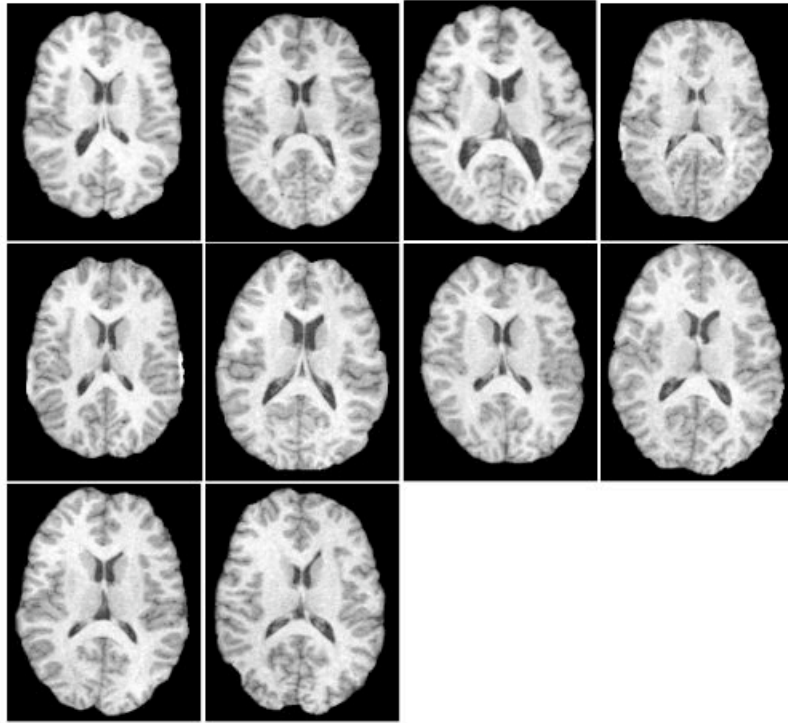


**Figure 2.1. The ten subject brain images used in these experiments. Variations in anatomical structures for these subjects can be observed.**

The program 3dFWHM in Analysis of Functional NeuroImages [AFNI, 37] based on algorithms described by Xiong [38] provides a good way to estimate smoothness of the mean brain images. In Xiong's method, the spatial correlation between voxels along each axis is characterized by FWHMx, FWHMy, FWHMz (FWHM, for "Full Width Half Maximum"). By definition, the reported filter width is the estimation of the FWHM of the Gaussian kernel needed to produce the current smoothness; so wider FWHM means a greater degree of smoothness and indicates more inter-subject misalignment during the registrations. Therefore the filter widths

resulting from 3dFWHM can be used as quantitative smoothness measurements. The 3dFWHM algorithm assumes an isotropic image of random or spatially uncorrelated variables. However a brain image has spatial correlation. Therefore, the smoothness estimation of the mean brain images from 3dFWHM includes the inherent spatial correlation of the anatomical images. The smoothness of the anatomical template colin27 was also estimated using 3dFWHM and the result was used as the inherent smoothness. The smoothness of the average brain images—generated using AIR, SPM, and the fully deformable model—were measured with 3dFWHM and compared to the inherent anatomical smoothness (from colin27).

### 2.2.2.3  FMRI effect-size (Experiment 3)

The registration accuracies from these three approaches were also evaluated by comparing the functional MRI signals (fMRI) acquired on the auto-segmented dorsal anterior cingulate region (dACC). One very consistent finding in fMRI studies of cognitive control is that the dorsal anterior cingulate region (dACC) shows significant activity when subjects make an error or have response conflicts while performing a task [39].  We used a recent fMRI study that we conducted on error and conflict processing [40] to examine how the different registration methods differ in their abilities to co-localize the functional MR signal across subjects. In this event-related fMRI study of implicit learning, subjects were asked to press a button as rapidly and accurately as possible based on the position of a stimuli on the screen.  Due to the time pressure of the task subjects made on average 20% errors.  As expected we found significant fMRI activation in the dACC on error versus correct trials. In the study, we compared the fidelity of dACC error signal when the dACC was identified using the three different methods: AIR, SPM, and the fully deformable registration.  The subjects who participated in the fMRI protocol included 8 of the 10 subjects whose structural images were compared in experiments 1 and 2.

The region of interest dACC (taken as the ROI described in Carter et al., [41]) was automatically segmented using the atlas-based segmentation technique (template->subject) with AIR, SPM, and the fully deformable model for each subject. Then, the signals were extracted on the segmented dACC after the functional images were aligned to the anatomical images. The fMRI signals were averaged across the ROI to produce an average activation signal across dACC for each trial. Voxel outliers were corrected using the criteria of one standard deviation, and the signal was normalized by the first time point at each trial for each voxel. For each subject the average correct activation timeseries was generated by averaging the activation timeseries across the correct trials and the average error activation timeseries was generated by averaging the activation time series across the error trials (i.e., when the subject inadvertently pressed the incorrect key). The peak activation differences between correct trials and error trials for all the subjects were statistically examined using a paired two-tailed t-test.

## 2.3    PEFORMANCE COMPARISON RESULTS

The performance of the fully deformable registration method is quantitatively evaluated in three applications including atlas-based segmentation, image normalization and fMRI signal localization. As predicted the fully deformable registration produced better results in all the three aspects.

### 2.3.1   Performance in Atlas-Based Segmentation (Experiment 1)

The mean overlap ratios across all the subjects for the right hippocampus and the ACC are shown in Figure 2.2. For both regions, the fully deformable model gave a higher mean overlap

ratio than AIR or SPM: 7.3% higher for right ACC and 15.6% higher for right hippocampus. A paired two-tailed t-test of the overlap ratios of the deformable model versus SPM was highly significant at $t(9)$ = -5.182, $p$ = 0.00058 (right ACC) and $t(9)$ = -6.372, $p$ = 0.00013 (right hippocampus). Similarly the t-test of the overlap ratios of the deformable model against AIR was significant at t(9) = -3.819, p = 0.0041 (right ACC) and t(9) = -3.8782, p = 0.0037(right hippocampus). However, there was no significant difference in mean overlap ratios between AIR and SPM at t(9) = 0.0494, p = 0.962 (right ACC) and t(9) = 0.1853, p = 0.857 (right hippocampus).
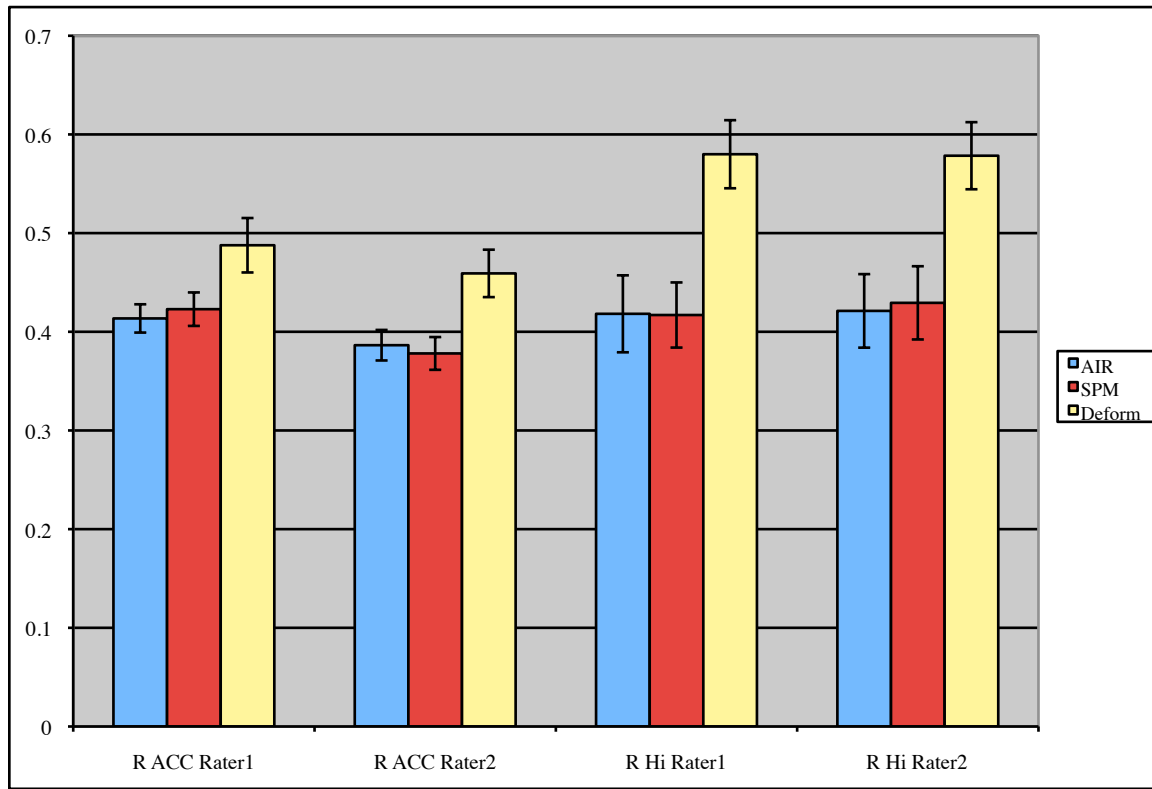


**Figure 2.2. Mean overlap ratios of right anterior cingulate cortex(R ACC) and right hippocampus (R Hi) across the 10 subjects were generated from atlas-based segmentation with AIR, SPM and the full deformable model (Deform).**

### 2.3.2 Performance in Image Normalization (Experiment 2)

For visual inspection the average brain images from AIR, SPM and the fully deformable model are shown in Figure 2.3; the template colin27 is also shown for comparison. As can be seen in Figure 2.3, the fully deformable model produces a much sharper average brain image with very clear boundaries, from which we can clearly identify the cortical sulci and subcortical regions.
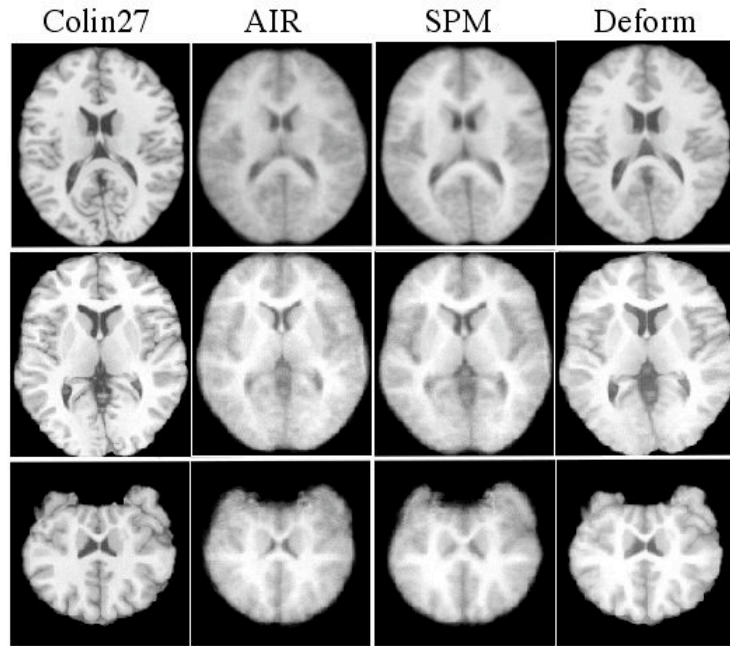


**Figure 2.3. The average brain from the 10 subject images using AIR, SPM and the fully deformable model. The template colin27 is also displayed for comparison.**

The smoothness measurements from 3dFWHM for the average brain images are shown in Table 2.1. As shown in the table, for the average brain image from the fully deformable model, the filter widths along three principal axes were very similar to the 3dFWHM results on the template colin27. Additionally, the filter widths were significantly smaller in all three dimensions than measurements of mean brain images from AIR or SPM. These results indicate that the fully deformable model introduces fewer inter-subject registration errors than AIR or SPM. The second experiment thus shows the superior performance of the fully deformable

method over AIR or SPM, at aligning the individual images to the standard MNI reference image or image normalization.

**Table 2.1. The Smoothness measurements of the averaged warped images**

| Methods\Axes | FWHMx | FWHMy | FWHMz |
|---|---|---|---|
| Colin27(template) | 5.31 | 5.77 | 5.59 |
| AIR | 16.65 | 19.44 | 17.31 |
| SPM | 14.62 | 17.66 | 15.74 |
| Fully Deform. | 6.96 | 7.68 | 7.60 |

The smoothness of averaged warped images from AIR, SPM, the fully deformable model was measured by the full width half maximum of the Gaussian smoothing filter along x, y, z axes through 3dFWHM. The smoothness of the template is also measure as a comparison, which described the inherent spatial correlation of the template colin27.

### 2.3.3 Performance in fMRI signal localization (Experiment 3)

As predicted, there was a greater fMRI signal on the error trials than the correct trials regardless of how the region was segmented (i.e., AIR, SPM, or the fully deformable model). However, as can be seen in Figure 2.4, with the fully deformable model we extracted higher group difference signals (average signals from error trials – average signals from correct trials) than AIR or SPM. As shown in Table 2.2, the paired two-tailed t-tests of the 8 subjects' peak signals of average timeseries for correct trials against for error trials was significant at $p = 0.0034$ for the fully deformable model, at $p = 0.0295$ for AIR, and at $p = 0.0668$ for SPM. Therefore, the t-test results were more significant with the fully deformable method, suggesting this method to be a more reliable extraction of the functional imaging signal.
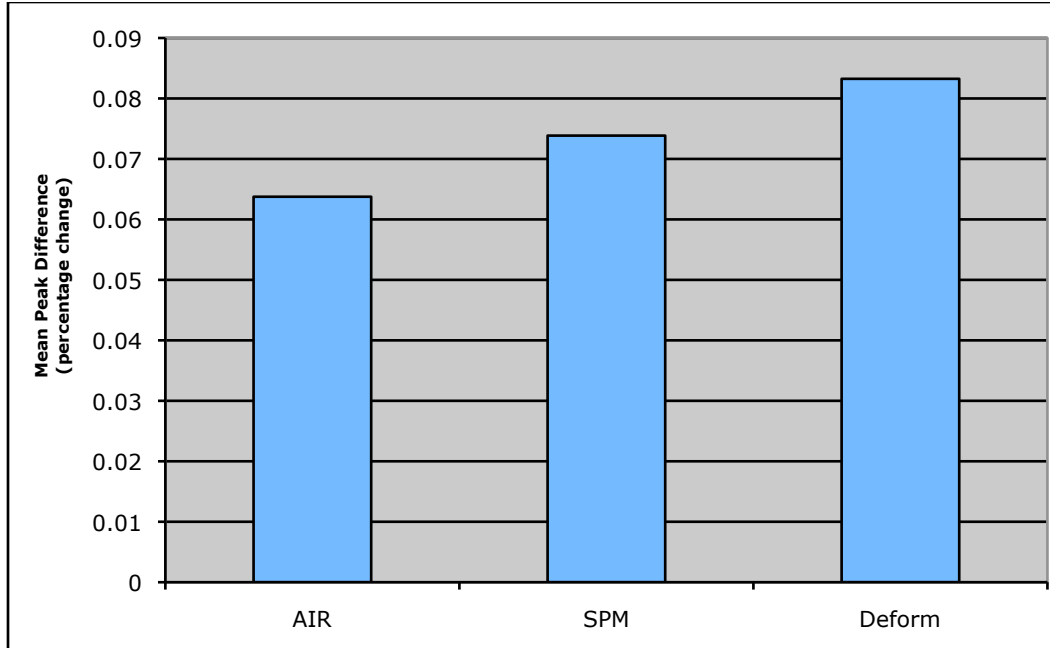
**Figure 2.4. An average of the peak fMRI activation difference across the 8 subjects. For each subject, peak activation difference is the percent signal changes on correct trials versus error trials.**

**Table 2.2. Statistical results on fMRI signals.**

| Method | t(7) | p value |
|---|---|---|
| AIR | -2.7269 | 0.0295 |
| SPM | -2.1681 | 0.0668 |
| Fully Deform. | -4.3499 | 0.0034 |

The paired two-tailed t-test results on peak signal of correct timeseries against peak signal of error timeseries based on the extracted ROI using AIR, SPM, and the fully deformable model.

## 2.4   SUMMARY

In this chapter, we quantitatively evaluated the co-registration performance of AIR, SPM, and the fully deformable model in atlas-based segmentation, image normlization and fMRI signal localization through a series of experiments. Compared with AIR and SPM, the fully deformable model produced significantly higher overlap ratios for the right hippocampus and the right anterior cingulate cortex in experiment 1, which demonstrates it can identify ROI more accurately than AIR or SPM. This leads to a more accurate co-localization of the ROI for functional images, thus it consistently produces more reliable functional signals as indicated in experiment 3.   Experiments 1 and 3 are based on template->subject co-registration, and experiment 2 is based on subject->template co (image normalization)-registration. In both situations (atlas-based segmentation and image normalization), the fully deformable model shows better performance compared to AIR and SPM in co-registration.

The fully deformable registration, however, is computationally expensive. On a G5 dual-processor Macintosh it took approximately 1 hour per brain.  This compared to approximately 10 minutes per brain for SPM (on an IRIX 64), and 2 hours for AIR on an IRIX 64.

Also, it should be noted that the registration methods are sensitive to accurate skull stripping. In this study we used BET [35], which was adequate for the 10 subjects in this sample. However, in other studies we have found some discrepancies that have led to certain registration inaccuracies. Recently, we have improved the stripping using an automated morphological method [42]. This study is limited by a relatively small sample, 10 subjects for the first 2 experiments, and 8 for experiment 3. Nevertheless the results for all three experiments were significant.

Overall, our results show that the fully deformable registration, compared to SPM and AIR, can improve anatomic alignment of brain images. Moreover, the improved registration from this method seems to lead to a more reliable mean BOLD fMRI signal. Currently, standard fMRI analysis pathways use AIR or SPM. Our results suggest that the full deformable could improve the reliability of the co-localized fMRI results. However, this comes at a cost of increased complexity of the registration and computation time. Moreover, replication of these results in a larger sample is also needed.

# 3.0    OPTIMUM TEMPLATE SELECTION FOR ATLAS-BASED SEGMENTATION

In the previous chapter, we found lower segmentation accuracy for the anterior cingulate cortex (ACC) than for the hippocampus in atlas-based segmentation [43]. We suspected that the lower accuracy for the ACC was due to the inter-subject variability in the gyral folding pattern in this structure. It has been estimated that approximately 30-60% of the population have a paracingulate sulcus (PCS) [17], a normal variant of the ACC in which there is an additional gyral fold. However, a single brain template Colin27 is unable to represent all the possible normal anatomical variations in the ACC, and thus the performance of atlas-based segmentation suffers.

## 3.1    TRADITIONAL TEMPLATE STRATEGIES IN ATLAS-BASED SEGMENTATION

Currently three different template selection strategies have been used in atlas-based segmentation: single brain atlas, probabilistic map and multi-atlas strategy. Most previous studies involving atlas-based segmentation have used a single fixed template strategy [44-46]. Widely used single-brain atlases include the Talairach template [22] and the standard MNI (Montreal Neurological Institute) template brain Colin27 [15]. The Talairach template was based on the brain of a 60-year-old female subject, and the MNI Colin27 template was averaged from

27 co-registered high-resolution scan of a young male subject. Since both atlases are based on one particular brain, neither one can fully represent the variety of anatomical structures present in the population of normal brains.

An alternative approach is to construct a probabilistic map by averaging brain images from a population of subjects after co-registration with a linear or nonlinear spatial transformation model to a stereotaxic space (individual space->stereotaxic space) [47, 48], such as MNI305 [49], ICBM152 and ICBM452 (ICBM: International Consortium for Brain Mapping)[50]. The probabilistic atlas map reflects the inter-subject variability within the population used for template construction. However, due to misalignment from registration errors or structural variabilities in the population, the structures in the population-based atlases are usually blurry and the atlases also lack anatomical detail.

Besides the above strategies, the nature of atlas-based segmentation (atlas->subject registration) also allows for the possibility of using multiple atlases. Multi-atlas segmentation has been found to improve the segmentation performance compared to the individual atlas and the averaged atlas approaches [51-55]. In the multi-atlas method, the segmentation procedure with each atlas is viewed as an independent classifier, and segmentation results from multiple atlases or classifiers are combined or fused to reach a consensus segmentation. Different fusion strategies have been explored including a simple voting rule with equal weight of training data [51, 53], a voting rule with equal weight within top ranked (similarity based sorting) atlases [55], or a voting rule with individual atlas weighted by their EM-based (expectation-maximization) performance [52]. The multi-atlas methods can substantially improve the segmentation accuracy [52].

In this section, we propose a novel optimum template selection strategy for atlas-based segmentation. Instead of choosing a fixed atlas such as Colin27 or a blurred population-based atlas (e.g., MNI305), we use a family of brain templates, and for each subject and each region of interest we choose the 'best' template. The intuition is that the variations in normal brain anatomy can be better represented as a small number of prototype atlases (e.g., for ACC, presence or absence of paracingulate) rather than as a single average brain. For each subject the template, which gives the optimum localized registration for a specific region of interest (maximum local similarity over a given ROI), is chosen as the optimum template for the segmentation. This approach has previously been shown to be effective in atlas-based segmentation of bee brain images [56]. In this study, this atlas selection technique was tested on two different human brain image data sets to segment multiple regions of interest (ROIs) including right anterior cingulate cortex (ACC), left and right amygdala, caudate, hippocampus, pallidum, putamen, and thalamus proper. For both data sets, the ROIs segmented using the optimum template selection method were compared to manual segmentations (ground truth) and the automated segmentations using a standard single template method.

## 3.2    OPTIMUM TEMPLATE SELECTION ALGORITHM

### 3.2.1   Materials

#### 3.2.1.1 Subjects and Brain Templates

Manual segmentation of multiple ROIs on two sets of data serves as the family of atlases for the template selection. Both data sets have been previously described in greater detail: data set 1 in

Chapter 2 and data set 2 at http://www.cma.mgh.harvard.edu/ibsr/. Brief descriptions are as followed.

Data Set 1: Nine subjects (6 male; mean age 24.3, range 20-32 years old; right-handed) participated. Scanning was done on a 1.5T GE CVi scanner with 3D SPGR (TR/TE = 5/25 ms, flip angle = 40º, FOV = 24×18 cm, slice thickness = 1.5 mm, matrix size = 256×192).

Data Set 2: Thirteen T1-weighted MR brain images (256×256×128) from Internet Brain Segmentation Repository (IBSR) and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital, available at http://www.cma.mgh.harvard.edu/ibsr/. Brain images were acquired from healthy control subjects.

Manual segmentations of the right ACC were done on data set 1, and multiple ROIs including the left and right amygdala, caudate, hippocampus, pallidum, putamen, and thalamus proper were manually segmented on data set 2. The manual segmentations of both data sets served as gold standard segmentations to evaluate the automated atlas-based segmentation results. The manual segmentations of the subjects also served as atlases on the subject when used as the template.

All the subjects in data set 1 were used as template candidates in the template selection algorithm. Every subject in data set 1 was warped to the 8 remaining atlases and the atlas with the best local registration accuracy was chosen as the optimum choice out of these 8 atlases. For data set 2, the atlas was selected from a randomly chosen 9-subject subset.

### 3.2.2 Template Selection Algorithm

For a given subject $S_i$, the steps for the optimum template selection from a family of templates $T = \{T_1, T_2, \ldots T_9\}$ are shown in Figure 3.1. The number of templates in this testing is fixed at 9 for both data sets.



**Template Selection Flowchart**

Register each template $T_k$ to the subject $S$, ($T_k$->S)

For each $T_k$->S, generated a warped image $W_k$ and ROI $R_k$

Generate a conjunct image C from all the $R_k$

Mask and crop $W_k$ and S at local area C respectively
Generate small region images $C_{wk}$ and $C_s$

Calculate the NMI between $C_{wk}$ and $C_s$ for each template

Select the template which gives the max NMI at local area

**Figure 3.1. Template Selection Flowchart. The processing steps that constitute the template selection model, which is used to choose the optimum template from a family of templates ($T_k$) for the segmentation of ROI (R) on a subject (S).**

In atlas-based segmentation, each template $T_k (1 \leq \forall k \leq 9)$ is registered to a given subject S using the deformable model ($T_k$ ->S), which creates a corresponding warped image $W_k$ and automated segmented ROI $R_k$. The registration accuracies of different templates for the subject S and the selected ROI, are evaluated at a common local region area $R$. The common local region $R$ is formed as the disjunction image of all the segmented ROIs from atlas-based

35

segmentation using the family of templates $T$. The common local area is defined as $R = R_1 \cup R_2 \cdots \cup R_n$, with $R_k$ as the segmented ROI using template $T_k$ in $T$ for a given subject.

To evaluate the local registration performance in $T_k$->S registration, normalized mutual information (NMI) is used as the metric to measure the local image similarity between the warped image $W_k$ and the subject S at the common ROI $R$. The following set of equations show the derivation for NMI:

$$NMI(x,y) = (H(x) + H(y))/H(x,y)$$
$$H(x) = -\sum_{i=1}^{k} P(i_x)\log_2(P(i_x))$$
$$H(y) = -\sum_{i=1}^{k} P(i_y)\log_2(P(i_y))$$
$$H(x,y) = -\sum_{i_x=1}^{k_x} \sum_{i_y=1}^{k_y} P(i_x,i_y)\log_2(P(i_x,i_y))$$

where x is the cropped local ROI image of each warped image (each template $T_k$->target) and y is the cropped ROI image of the target S at the same local area $R$; H(x), H(y) are the entropies and H(x,y) as the joint entropy of x and y. NMI describes the similarity of the warped image and the target image at a local ROI area $R$, which also evaluates local registration accuracy and the performance of the template $T_k$ in the segmentation of the ROI. Using the above equations, the template that gives the maximum local NMI is chosen as the locally optimized template for each subject's ROI.

Atlas-based segmentation labels the anatomical regions on individual images by registering the atlas brain image to the individual brain image space. We refer to the registration method we use as the Automated Labeling Procedure (ALP). It is derived from the methods used by Chen [57], and consists of a series of preprocessing steps (including skull stripping and cropping) and a series of registration techniques including hierarchical registration [58] and

demons based registration [59]. We have implemented this method using the Insight Segmentation and Registration Toolkit [ITK, 60]. Details of the method are described in Chapter 2.

## 3.3    ATLAS EVALUATION METHODS

### 3.3.1    Overlap Ratio

Overlap ratio (OR) was used to quantify the segmentation quality of each registration. Similar to Dice metric [61], OR is defined as the ratio of overlapping voxels to total voxels, as given below:

$$overlap\ ratio = \frac{vol(B \cap \hat{B})}{vol(B \cup \hat{B})}$$

where B is the ground-truth mask, and $\hat{B}$ is automatically segmented set. As shown in Figure 3.2, a perfect overlap between the manually segmented group-truth B and the automated $\hat{B}$ will lead to an OR = 1, while a smaller overlap will result in a smaller value of OR.

Ground-truth Mask $B$        Auto-segmented Set $\hat{B}$

Vol I     Vol II     Vol III

$$Vol\ II = B \cap \hat{B}$$
$$Vol\ I = B - B \cap \hat{B}$$
$$Vol\ III = \hat{B} - B \cap \hat{B}$$
Overlap ratio = Vol II / (Vol I + Vol II + Vol III)
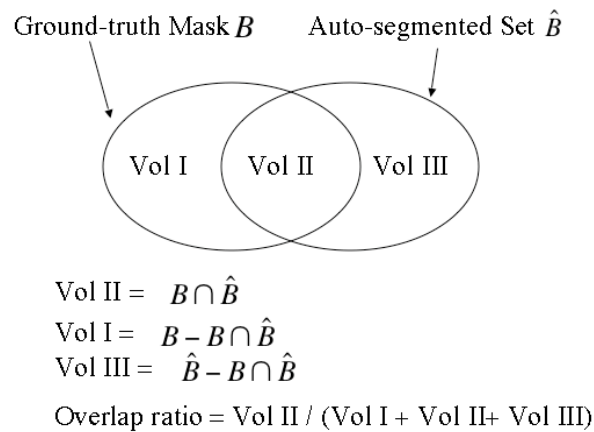
**Figure 3.2 The illustration of the overlap ratio.**

The performance of the optimum template was measured as the mean overlap across multiple subjects for the ROI segmentation.

### 3.3.2 Volume Consistency

Volume consistency between the automatically segmented ROI and the hand drawn region was evaluated using the intraclass correlation coefficient (ICC), with a two way mixed and absolute agreement model. The reliability of the estimated volume from the optimum template method was compared to the average reliability of individual templates for each ROI.

### 3.3.3 Single Atlas Strategy

For each data set, the segmentation performance of the individual atlas was evaluated based on a leave-one-out approach. Each of the 9 subjects for data set 1 was chosen to serve as an individual template and was registered to the remaining 8 subjects; 9 subjects out of the 13 subjects for data set 2 were randomly chosen as individual atlases and registered to the remaining 12 subjects. For each ROI, the performance of a single atlas was evaluated by the mean overlap ratio between the segmented ROI and the manual ground truth across multiple subjects using the same atlas. The average performance of the single template strategy was measured as the average overlap ratio across multiple templates. Additionally, the standard MNI (Montreal Neurological Institute) brain colin27 [15], which carries high anatomical details and has a high spatial resolution ($1mm^3$ voxel size), was also used as the template to segment right ACC on data set 1 for comparison.

## 3.4    RESULT: IMPROVED OVERLAP RATIO AND VOLUME CONSISTENCY

The segmentation evaluation results for both methods (individual template and optimum template) were compared for each data set. Both methods used exactly the same pathway for the registration as well as the same thresholds to remove the edges of the automated segmentations. The only difference was that they used different atlas selection strategies. We found that for most of the ROIs the optimum template produced significantly and consistently better segmentation results compared to the single template method.

### 3.4.1    Data Set 1

As predicted, in data set 1 the template selection method produced significantly better mean ORs for than any single template in the atlas-based segmentation of right ACC. The mean ACC OR when registering with a single template ranged from 42.7% to 52.7% (mean OR 49.5%), and the mean OR using the standard template MNI Colin27 was 47.3%. The mean OR with the optimum template selection method was 54.7% (using 8 templates) and the OR reached 57.7% when an optimized subset of templates was used (3 templates, Figure 3.3). As shown in Figure 3.3, the performance of the optimum template selection method was better than that of any template candidate or the standard MNI template Colin27. A two-tailed paired $t$ test also indicated that the registration result from optimum template (8-template case) based method was significantly better than the results from single templates at $t(8) = 4.353$, $p = 0.0024$.
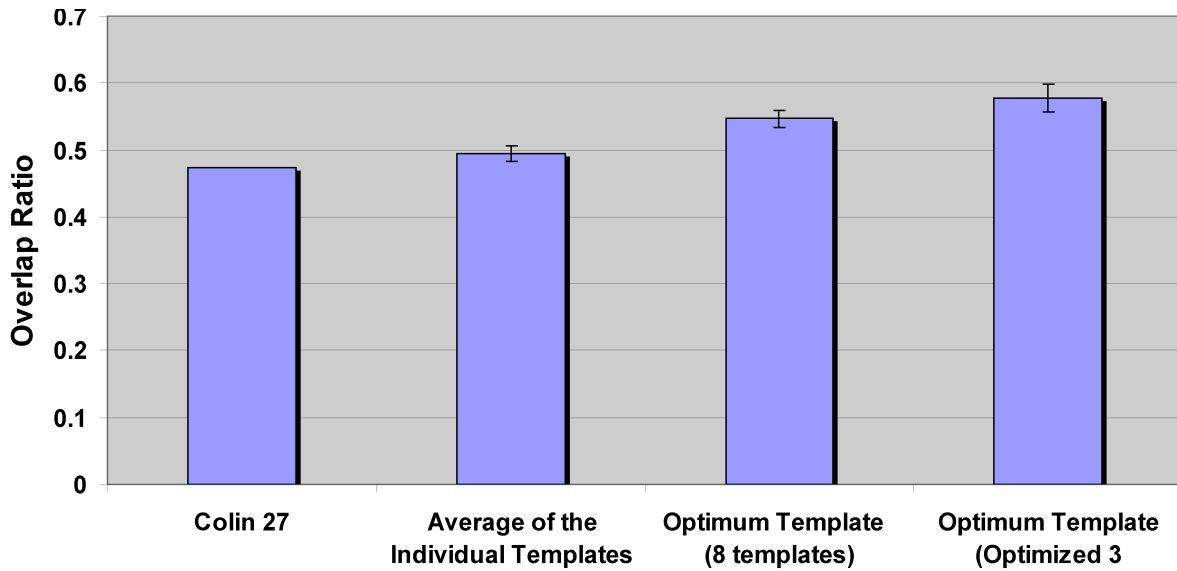
**Figure 3.3. In the anatomical classification of right ACC, mean overlap ratios (OR) are compared between individual template strategy (average of the mean OR over the 9 templates) and the optimal template method with 8 templates or with the optimized 3 templates. Error bars were calculated as standard error of the difference between the mean OR of the optimal template and the individual templates. The OR for the Colin template is shown for comparison.**

### 3.4.2 Data Set 2

The automated anatomical labeling results for multiple regions (left and right amygdala, caudate, hippocampus, pallidum, putamen and thalamus proper) on the IBSR dataset (data set 2) using the template selection method and the single template method were each compared to manual tracings using OR and volume agreement. For most of the ROIs, the template selection method consistently provided more reliable region classification than the single template method. The mean percent ORs and ICCs between the estimated volumes and manual traced volumes for both methods are shown in Figure 3.4.

The template selection method gave a higher mean OR than the single template method for all ROIs. The OR percentage increase in the template selection method ranged from 4.4% to

13.1% (mean increase 8%). The differences in ORs between the two methods were highly significant for all ROIs using the pair wise t-test (for right amygdala, $t(8) > 3.175$, $p < 0.013$; for the remaining ROIs, $t(8) > 4.36$, $p < 0.002$).

Additionally, the template selection method provided a more reliable volume estimate. The ICCs between the estimated volumes and the manual tracings from the template selection method were higher than the results from single template for 11 of 12 ROIs. The changes in the ORs were consistent with the changes in ICCs from the volume agreement. For example, the large increase of ORs in regions like the left caudate (11.1% increase), right caudate (12.0%), and right pallidum (13.1%) corresponded to more improved ICCs in the left caudate (ICC 0.95), right caudate (ICC 0.93) and right pallidum (ICC 0.91) respectively. For all of the ROIs except for left and right amygdala and right putamen, the template selection method produced very reliable ICCs of volume estimates and the results are comparable to ICCs of inter-rater manual tracings. These findings are essential to MR volumetric studies on such ROIs. For instance, the template selection method produced ICCs of 0.95 for the left caudate and 0.93 for the right caudate, while previous research has reported inter-rater reliability of 0.94 for the left caudate and 0.95 for the right caudate [62].
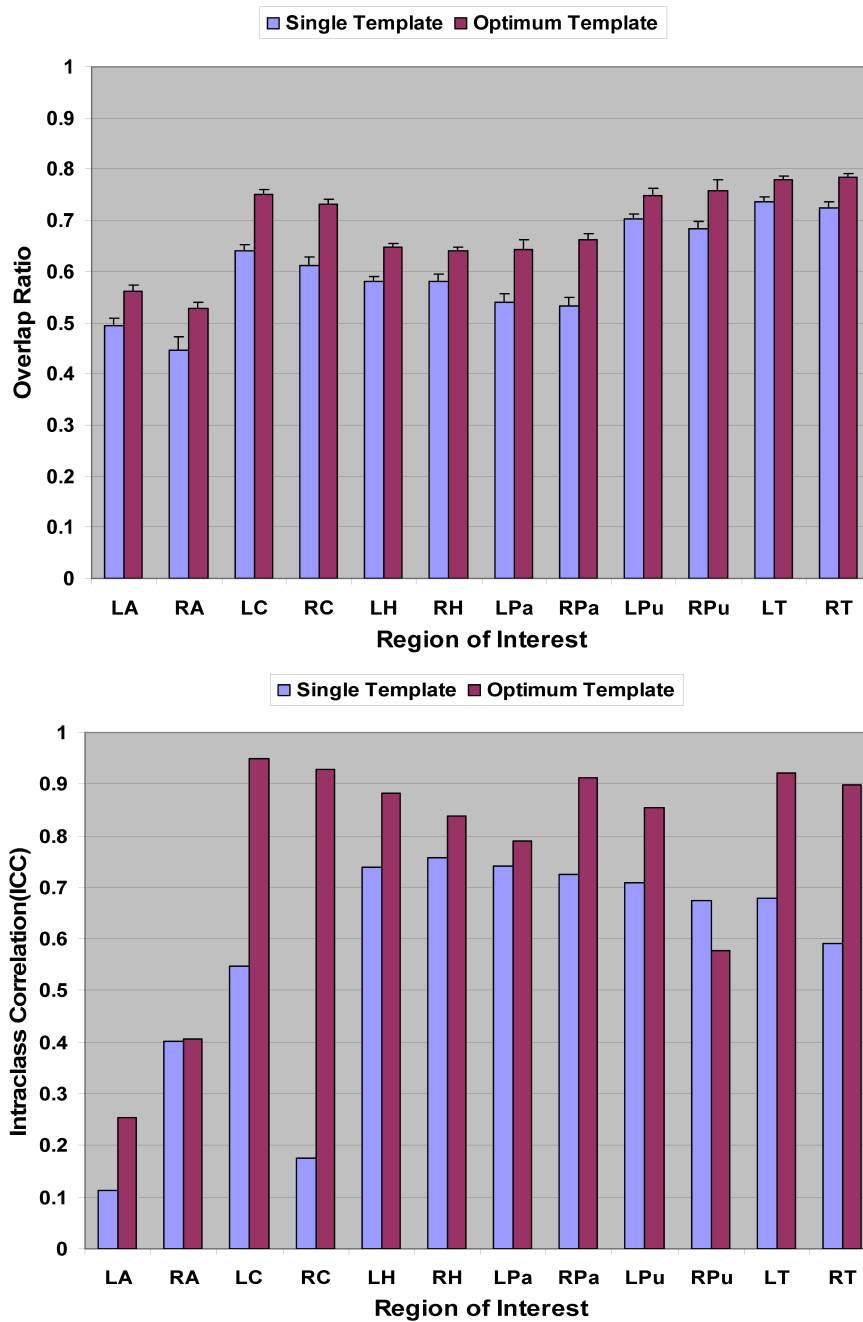
**Figure 3.4. Comparison of the reliability of the automated ROIs using the template selection method and using a single template. Key: LA—left amygdala, RA—right amygdala, LC—left caudate, RC—right caudate, LH—left hippocampus, RH—right hippocampus, LPa—left pallidum, RPa—right pallidum, LPu—left putamen, RPu—right putamen, LT—left thalamus proper, RT—right thalamus proper. Top: mean percent OR comparion. Bottom: the intraclass correlation coefficients (ICCs) of volume agreement for both methods.**

To illustrate the strong agreement between the estimated volumes and the manual volumes, the estimated volumes of four ROIs from data set 2 were plotted for both methods against the manual volumes in Figure 3.5. As Figure 3.5 illustrates, the accuracy of volume estimates improved considerably by using the optimum template method. Overall, the optimum template method provided better volume estimates than the single template method. However, for the left and right amygdala, neither method gave a good estimate. This is because the registration failed to provide an accurate region classification with any of the template candidates.
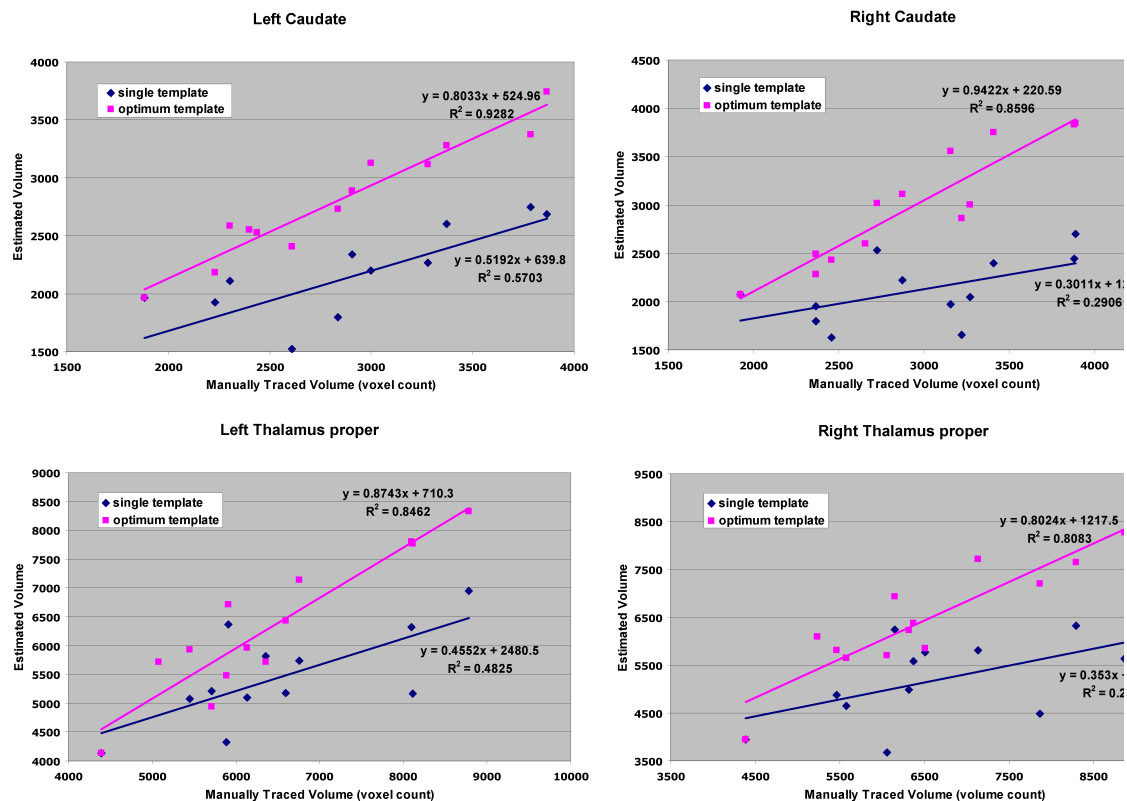


**Figure 3.5. Region classification performance of the optimum template method and the single template method. The absolute voxel numbers of segmented ROIs from both methods are compared to manual segmentation respectively using the linear regression model. Four ROIs (left and right caudate, and left and right thalamus proper) were analyzed for 13 subjects. Subject 6 was used as the atlas in the single template method; the optimum template was chosen from a 9-subject subset.**

## 3.5    TEMPLATE NUMBER AND PERFORMANCE

In the previous test on Data Set 1, for each subject the remaining 8 brain images were used as template candidates for the segmentation of the right ACC in the template selection method. However, adding more templates in the optimum template method leads to increased computation time, so it is important to determine how many templates are needed to achieve robust automated anatomical labeling of certain ROIs with the template selection method.

In order to test how the performance of the template selection method changes with the number of templates and decide how many templates are sufficient for the classification of specific ROI, we tried the template selection method on data set 1 using all subsets of the 8 different templates (from 1 to 8 templates in each subset). The performance of the template selection method with a particular number of template candidates was estimated as the average overlap ratio (OR) across all subjects for all subsets of templates with that cardinality. For example, the performance of the template selection method with 2 templates was estimated as the mean OR across the 7 subjects (excluding the 2 templates) and across all possible template combinations ($C_9^2 = 36$).

The mean ORs are plotted in Figure 3.6 against the number of templates used. As expected, as the number of templates increased, the average performance of the template selection method improved. For each subset of templates (N templates), the performance was estimated by the average OR across the remaining subjects (9-N). We observed that among the template subsets of the same cardinality, the template set with the widest anatomical variations

performed better than other combinations of templates. For example, in the automated classification of right ACC, we discovered there was a sufficient set of three templates, which included three prototypes: one template with a paracingulate sulcus, one with a thick anterior cingulate cortex, and the third with a thin anterior cingulate cortex. The template selection method with this optimized subset had the best performance and achieved an average OR of 0.577 (similar to that of the 8 template case), as shown in Figure 3.7. This suggests that in addition to template number, template variability is also important, such that with an appropriately variable set of templates, a fewer number of templates are sufficient for high accuracy.
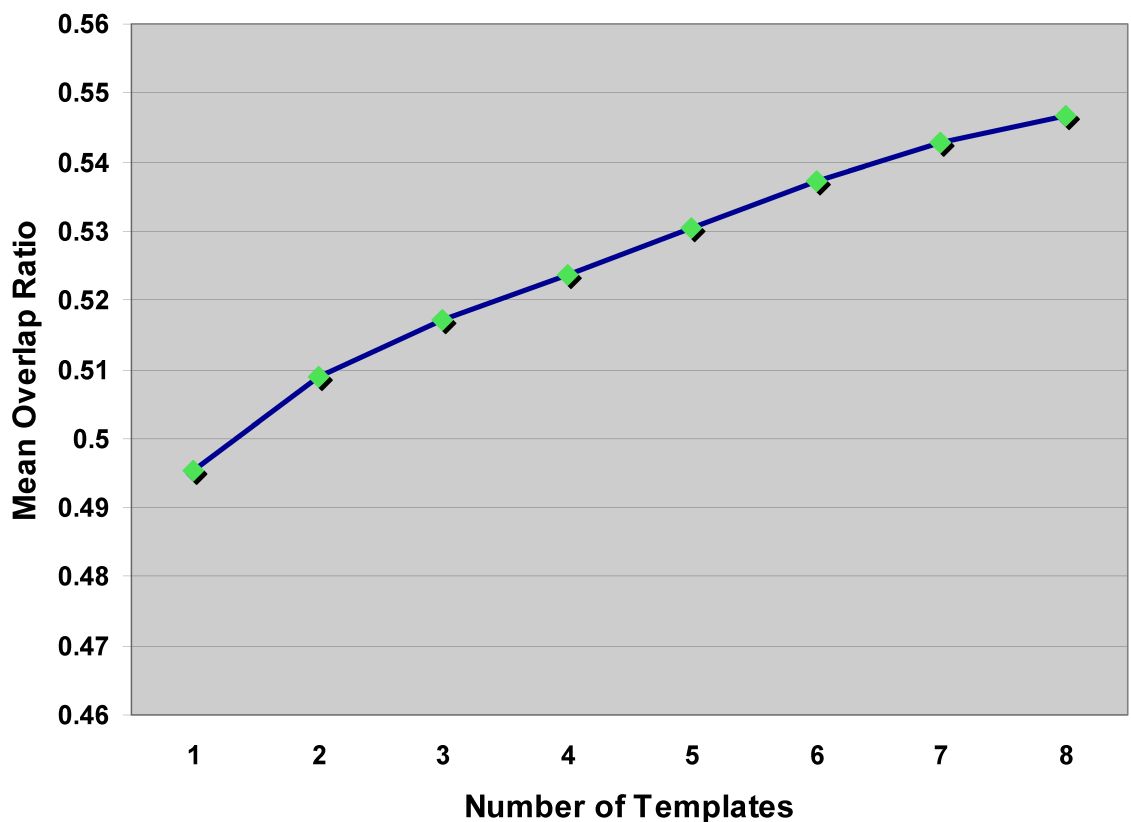


**Figure 3.6. The performance of the multiple template method with different number of templates. The mean ORs across all the combinations of same-number templates were plotted against the number of templates used.**

For one subject image in data set 1, the sufficient set of 3 template images and the corresponding NMI results from the template selection model are shown in Figure 3.7. It can be noted from the figure that the target subject has a paracingulate sulcus (in red), and that template 2, which has a similar paracingulate sulcus (in red) was automatically selected as the optimum template to segment the right ACC on the subject using the maximum NMI between the warped image and the target subject image at local ACC areas. The OR of the automatically labeled right ACC from the 3 templates against the manual tracing on the target image was also calculated and compared to validate the performance of the template selection model.

| Target Subject | Template1 | Template2 | Template3 |
|---|---|---|---|



|  | NMI 1.337 | NMI 1.357 | NMI 1.330 |
|  | OR 38.1% | OR 57.9% | OR 35.1% |
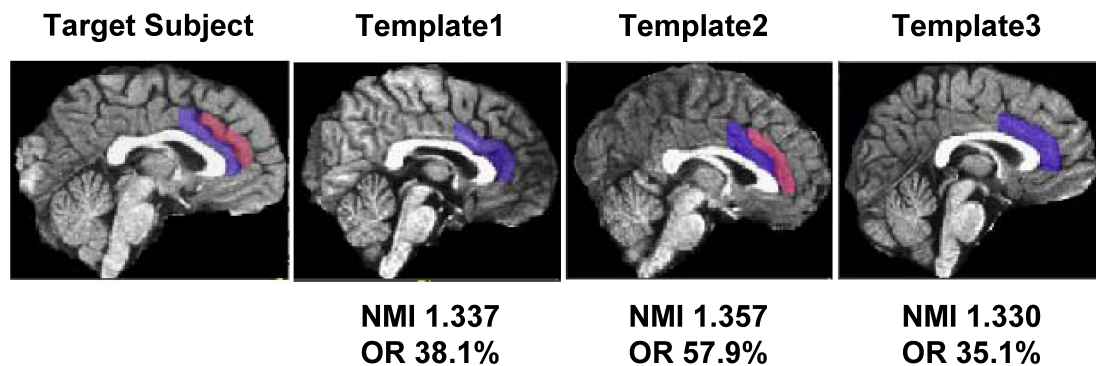
**Figure 3.7. For data set 1 the optimum template selection model selects the best template from a family of 8 templates to segment the right ACC on the target image. Three of the templates and the target subject images are shown here; the hand-drawn ACC on the subject and templates are also displayed in color (cingulate in blue and paracingulate in red). Also shown are the normalized mutual information (NMI) calculated by comparing the warped template with the target image, and the overlap ratios (ORs), calculated by comparing the automated segmentation with the ground truth manual segmentation.**

### 3.6    SUMMARY

In this section, multiple prototype atlases were used to address the normal brain anatomic variations in the atlas-based segmentation of MR brain images.

46

The template selection algorithm uses normalized mutual information to choose the template (from a family of templates) that gives the best local registration accuracy. This template selection model is of special use for those regions with high variability across subjects such as cortical structures [63], where a single template can not readily capture the variability. In this, the template selection model produced significantly better ORs and more reliable volume estimations in the analyzed multiple ROIs than the single template strategy. In the template selection method, the segmentation accuracy improves as the number of the templates used increases. In addition to the template number, the anatomical variability within the templates is also important, such that a fewer number of templates with appropriate anatomical variations are sufficient to achieve high accuracy in the atlas-based segmentation. For example, in the case of right ACC, we found a set of 3 prototype templates with wide anatomical variations, which performed better than other template combinations.

Nevertheless, this improved registration accuracy with the template selection model is achieved at the cost of higher computation load. Multiple non-rigid registrations are required in order to evaluate the performance of multiple templates. Also, although this method produced improved anatomical classification accuracy for all the analyzed ROIs, it did not give satisfactory region estimates for the left and right amygdala. Alternative classification methods (perhaps using feature-based registration) should be used to improve the automated labeling of the small and difficult to segment regions such as the left and right amygdala. A more advanced method may be needed to evaluate the performance of the templates at such regions.

The number of available atlases limits the registration accuracy. More atlas candidates may lead to higher registration accuracy, but with extra computation load, since we need to register each template to the target subject in order to evaluate the templates. Also the number of

atlases needed is related to the normal anatomic variations of the region to be segmented. Different ROIs may require different numbers of atlas prototypes for the automated anatomical classification. In this study, we discussed the possible atlas prototypes for the anterior cingulate cortex (thin ACC, thick ACC, and with a paracingulate sulcus). Further research is needed to explore the normal variations of the brain anatomy.

Multiple templates are needed in the template selection method. The templates can either be manually traced locally by experts, or downloaded from a public database. There are many manually labeled atlases available online, such as the IBSR dataset used in this paper (http://www.cma.mgh.harvard.edu/ibsr/), which consists of 18 high-resolution T1-weighted MR image data with expert segmentations of 43 individual structures.

This method chooses the best atlas from a family of atlases for each subject and ROI, and the optimum template selection method is independent of the registration techniques. In our study, we used the deformable automated labeling pathway for the inter-subject registration. The optimum template selection method can be easily accommodated into alternate available pathways such as Automated Image Registration (AIR) or Statistical Parametric Mapping (SPM).

# 4.0    TEMPLATE SELECTION IN IMAGE NORMALIZATION

The accurate alignment of brain MR images is essential for most neuroimage analyses pipelines. The most common approach relies on the alignment of individual brain images into a common template space [64]. However, individual brains in a population do not conform to a unique anatomic morphology variant. The normal and pathological variations in human brain anatomical structures present a fundamental problem for using a single template in the image alignment [47, 65]. For geriatric populations the accurate alignment to the selected template can be particularly problematic due to increased variability in brain morphology, which is due in part to age-related atrophy [66, 67].

Chapter 3.0   described and evaluated the optimum template algorithm in atlas-based segmentation, in which the best template is selected from a family of templates for a given subject and ROI to improve the segmentation accuracy. The algorithm may select different templates for different subjects or ROIs, and thus it is not suitable for image normalization, as the goal of image normalization is to warp individual brain images into a common template space.

This chapter will first discuss the available template selection strategies in image normalization, and then will propose and develop a novel multi-template strategy applicable to the image normalization of the elderly population. In this method, we use multiple templates as intermediate prototype brain images in the image alignment. The purpose of this strategy is to

characterize the brain anatomical structural variations in the elderly population while preserving brain structural details. We hypothesize that image normalization into Colin27 using intermediate images (multi-template strategies) will lead to improved registration accuracy.

Different strategies for choosing intermediate templates from a <u>training</u> dataset are proposed and their performances in improving registration accuracy of image normalization are evaluated with a <u>testing</u> dataset. The performance of the multi-template strategy to improve registration accuracy is assessed for different registration algorithms.

## 4.1    TRADITIONAL TEMPLATE STRATEGIES IN IMAGE NORMALIZATION

Similar template strategies have been explored in image normalization as in atlas-based segmentation, including single brain atlases and averaged brain atlases. Using an unrepresentative single brain template in the image normalization [e.g. the Talairach atlas [22], the MNI Colin27 [15]] introduces a bias in the image registration by favoring subjects similar to the template compared to subjects substantially different from the template. Population-based average templates, such as MNI305 [49] and ICBM452 [50], have been constructed to reduce the bias in the registration by keeping the group features. However, population-based average templates are blurred and lack anatomic details due to misalignment from registration error or structural variability in the population.

In addition to the above approaches, some template selection strategies have also been proposed specifically for image normalization. For example, an unbiased template was constructed by choosing the brain image closest to the mean geometry of the population under consideration [68, 69]. Other methods, such as implicit reference pair-wise (IRP) image

registration [70], a template-free approach [71], and an unbiased diffeomorphic atlas [72], jointly register a population of images into a hidden common space with different similarity cost functions and different deformation models.

An alternative approach, known as rapid alignment of brains by building intermediate templates (RABBIT), constructs an intermediate template from the target template for a given individual brain image via a statistical deformation model. The statistical deformation model is built by principal component analysis (PCA) on the deformation fields of the training dataset. The final deformation field (template→individual brain image) is a combination of two deformation fields (template→intermediate template, intermediate template→individual brain image) [73].

## 4.2  MULTI-TEMPLATE METHOD

In this proposed novel multi-template strategy, multiple prototype templates are used as intermediate brain images for the image alignment. Instead of registering an individual brain image directly to the standard MNI space Colin27 during image normalization, our method chooses intermediate prototype templates as a bridge between the individual brain image and Colin27 (subject→intermediate brain images→Colin27). The final deformation field for subject→Colin27 registration is the combination of two deformation fields from subject→intermediate brain images→Colin27 registration.

The multi-template method consists of three steps:

1) Identification of the multiple prototype templates that represent the structural variations in the elderly population (or the training database).

2) Selection of the appropriate intermediate brain images from the training data for each subject of the testing datasets. These intermediate brain images are used for image registration (subject→intermediate brain images).

 3) Transformation of the intermediate brain images into Colin27 (intermediate brain images→Colin27) followed by combination of the deformation fields from the transitional registrations (subject→intermediate brain images, intermediate brain images→Colin27) in order to achieve the deformation of subject brain images into Colin27 space.

This step-wise approach stems from the anatomic diversity of brain images. Images do not necessarily fit a single morphologic prototype, but rather can be better aligned using a set of intermediate prototypes.  To achieve this goal we create a set of template prototypes from the ADNI training dataset and then utilize these prototypes as intermediate registration targets when aligning subjects from a new study (the testing dataset).

### 4.2.1   Materials

In this study, two separate databases of MR brain images were used: one for generating the set of prototype templates (referred to as the <u>training</u> dataset) and the other for testing the performance of the multi-template method in image normalization (referred to as the <u>testing</u> dataset).

### 4.2.1.1 Training Data

Training data in this study was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-

profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) . Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

Mild cognitive impairment  (MCI) is defined as the presence of cognitive decline not warranting the diagnosis of dementia and/or evidence of decline over time on objective cognitive tasks but with preserved basic activities of daily living (Report of the International Working Group on MCI –Stockholm, 2004). The typical rate at which MCI patients' progress to AD is 10-15% per year [74, 75].

The ADNI dataset contains MR images collected as a part of a large multi-site neuroimaging study of aging. The dataset is available for use by the research community to support studies of aging and dementia. This dataset has been particularly useful for the current study because it contains high-quality T1-weighted brain MR images from a diverse sample of elderly individuals.

One hundred MR Brain images of healthy controls and MCI elderly subjects from the ADNI database were used as the training database for this study. The ADNI training data were gender and age-matched (50M/50F, mean age = 75.4±6.1). Demographic information for the selected data is included in Table 4.1.

Sagittal T1-weighted high-resolution brain images were acquired with 3D magnetization-prepared rapid acquisition with gradient echo (MP-RAGE) using a range of parameters: 160-208

slices, 192 x 192 (or 256 x 256) matrix, FOV = 240 x 240 (or 256 - 260x 240) mm$^2$, TR = 2400

(or 2300, 3000) ms, TI = 1000 (or 853-900) ms, Flip angle = 8°(or 9°), slice thickness = 1.2mm.

Detailed information of image acquisition in the ADNI study is described by Jack et al.[76].


### 4.2.1.2 Testing Data

The testing MR data were acquired locally as part of an ongoing study focusing on functional

MR imaging of elderly depressed individuals (major depression and subsyndromal depression).

Images from seventy elderly subjects were included. Detailed demographic information is

provided in Table 4.1.

Axial T1-weighted images were acquired with 3D MP-RAGE on a Siemens Trio 3T

scanner using a 12-channel head array coil and the following parameters:: 176 slices, 224 x 256

matrix, FOV = 224 x 256 mm$^2$, TR = 2300 ms, TE = 3.43 ms, TI = 900 ms, Flip angle = 9°, slice

thickness = 1mm, and no gap. This study was approved by the University of Pittsburgh

Institutional Review Board (IRB) and a written informed consent was obtained.

**Table 4.1 Demographic information of the training data and the testing data.**

|  | Training dataset | | | Testing dataset | | |
|---|---|---|---|---|---|---|
|  | Female | Male | All | Female | Male | All |
| Subjects (N) | 50 | 50 | 100 | 47 | 23 | 70 |
| Age, years | 76.0±5.8 | 74.8±6.4 | 75.4±6.1 | 70.7±8.2 | 69.8±7.3 | 70.4±7.8 |
| Healthy | 33 | 33 | 66 | 22 | 8 | 30 |
| *MCI | 17 | 17 | 34 | 2 | 1 | 3 |
| $^\triangle$Others | 0 | 0 | 0 | 23 | 14 | $^\triangle$37 |

* MCI - Mild cognitive impairment.

$^\triangle$ Of the 37 subjects, 28 with major depression, 6 had subsyndromal depression and 3 had mild AD.

## 4.2.2   Finding Prototype Templates

A bottom-up clustering algorithm is used on the <u>training</u> dataset to identify a set of intermediate prototype templates that spans the space of brain morphometry in the elderly population [23-26]. The procedure to identify multiple representative templates from the training data is illustrated in Figure 4.1. This procedure includes: a) partitioning the training data into subgroups using unsupervised classification, and b) assigning a representative template image for each subgroup.
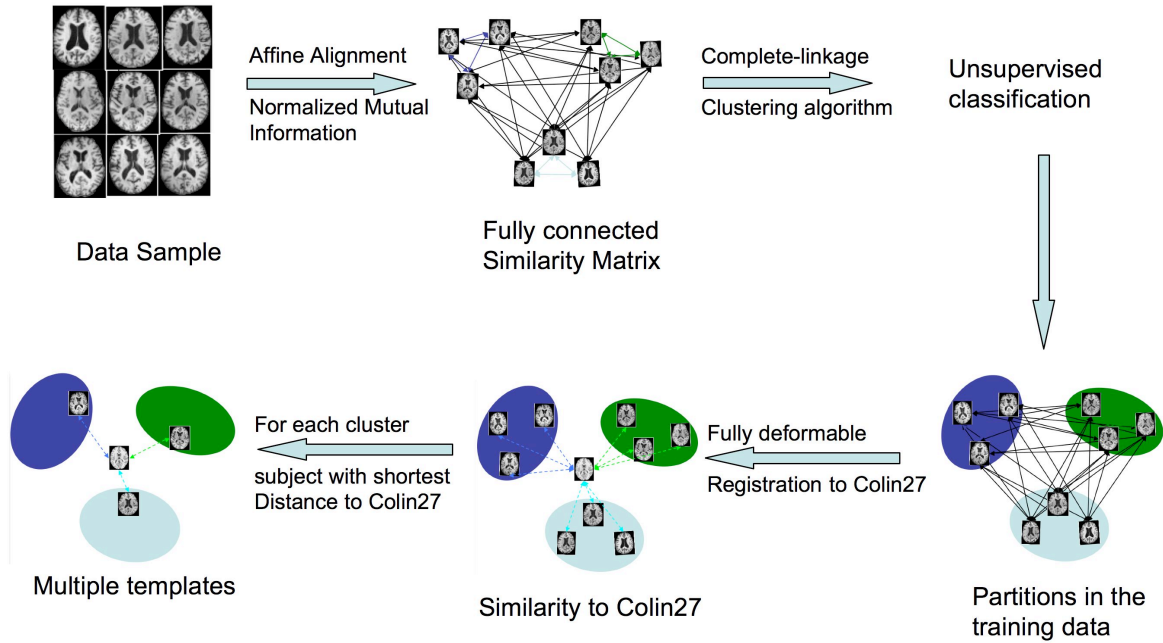


**Figure 4.1. The procedure to find prototype templates in the training data.**

The complete-linkage clustering algorithm in graph theory [23-26] was used to divide the training data into a finite number of clusters (subgroups) by progressively merging clusters with minimum distance or maximum similarity. The distance or similarity of any two images (in the training data) was evaluated by the normalized mutual information of the two co-registered images (affine co-registration). The normalized mutual information was calculated by the following equations:

$$NMI(x,y) = 2.0 * MI(x,y)/(H(x) + H(y))$$

$$MI(x,y) = H(x) + H(y) - H(x,y)$$

$$H(x) = -\sum_{i=1}^{k} P(I_x)\log_2(P(I_x))$$

$$H(y) = -\sum_{i=1}^{k} P(I_y)\log_2(P(I_y))$$

$$H(x,y) = -\sum_{i_x=1}^{k_x} \sum_{i_y=1}^{k_y} P(I_x,I_y)\log_2(P(I_x,I_y))$$

where $x$ and $y$ are any pair of co-registered images from the training data, and $I_x$ and $I_y$ stand for the intensity of image $x$ and $y$ respectively.

The affine co-registration corrects for differences in brain size and orientation of the images. A fully connected similarity matrix (100x100) $X_0$ is constructed after the pair-wise affine co-registration, where the element $X_0(i,j)$ is the similarity measurement between the co-registered training images $i$ and $j$. The similarity matrix $X_0$ is symmetrical and the values on the diagonal $X_0(i,i)$ are set to 0 in the clustering algorithm.

The clustering algorithm starts with the fully connected similarity matrix $X_0$, labels each brain image in the training dataset as a subgroup at the 0th iteration, and then iteratively merges the two closest or most similar subgroups (subgroup $m$ and $n, m > n$) into one big subgroup $n$ based on the minimum distance or maximum similarity. When subgroup $m$ and subgroup $n$ are merged at the $(t+1)th$ iteration, the similarity matrix $X_{t+1}(i,j)$ $\forall i,j < 100 - (t+1)$ is updated such that the similarity of any subgroup $i$ to $n$ $X_{t+1}(i,n)$ is the minimum of $X_t(i,n)$ and $X_t(i,m)$.

We did not know a priori how many clusters would most effectively divide the training dataset, and thus we chose to avoid setting a predetermined cluster number K as the stopping criteria for the complete-linkage algorithm. Instead, we used a similarity threshold (mean similarity -0.5*standard deviation) as the stopping criteria for the clustering algorithm. The mean

similarity and standard deviation are calculated from $X_0(i,j)$ $\forall i > j$ in the initial similarity matrix $X_0$. The flowchart of the clustering algorithm is shown in Figure 4.2.

The representative brain image for each cluster was then defined as the brain image in the cluster that is most similar to Colin27, since the ultimate goal of the image normalization task is to transform each subject's brain image into the standard MNI Colin27 space. The similarity between Colin27 and each brain image in the training database is measured by normalized mutual information after the deformable co-registration.
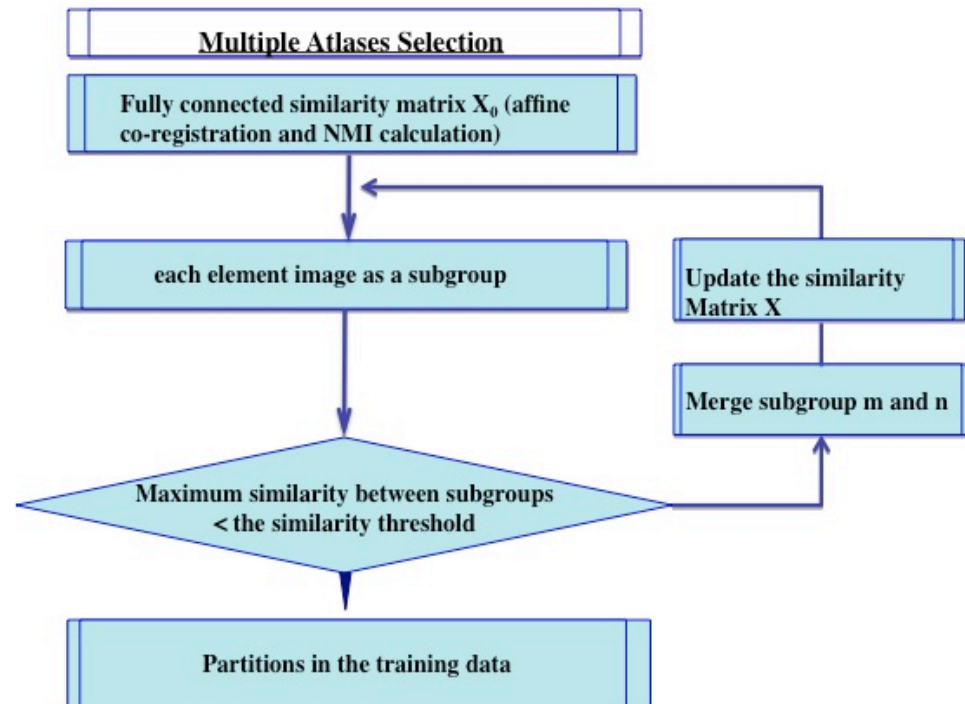


**Figure 4.2. The flowchart of the clustering algorithm on the training data.**

### 4.2.3 Image Normalization Using Multiple Templates

For any given subject from the testing dataset, the similarity between the brain images of the subject and any training subject is calculated as the normalized mutual information of the two images after affine co-registration. The brain image in the training dataset with maximum normalized mutual information is labeled as <u>the most similar image (s-image)</u> for this subject. The s-image has been classified into a subgroup of the training data using the complete-linkage clustering algorithm (as described in 4.2.2), and <u>the representative brain image</u> of this subgroup is chosen as <u>the representative or prototype image</u> (r-image) for this subject.

To normalize a testing brain image into the standard MNI Colin27 space, we explored three intermediate template selection strategies:

(1) Choosing the most similar brain image (s-image) from the training data as the only intermediate template, and then carrying out the registration as subject→s-image→Colin27.

(2) Choosing the representative image (r-image) of the subgroup that the s-image belongs to (subject→r-image→Colin27);

(3) Choosing both the most similar image and the representative image as intermediate templates to Colin27 (subject→s-image→r-image→Colin27).

### 4.2.4 Registration algorithm

We differentiate the registration between the training data and Colin27 (e.g. s-image→Colin27, r-image→Colin27, and s-image→r-image→Colin27) from the registration of the testing data to the selected intermediate images (subject→s-image or r-image).

58

The registrations between the training data (s-image→r-image) and MNI Colin27 (s-image→Colin27 and r-image→Colin27) are one-time only investments. These registrations can be computed offline, and the computed deformation fields can be repeatedly used for image normalization in future studies. Therefore we can afford to use computationally expensive algorithms or labor-consuming methods to improve the registration accuracy. In this study, for registrations we used a highly deformable registration algorithm with a large number of iterations (500 iterations) [43].

However, the registration of the testing data into the selected intermediate brain template(s) is a different challenge. The computed transformation field is specific for each subject in the study and cannot translate to other studies. This image normalization procedure is a time-consuming computational burden for each study, and needs to be fast and fairly accurate. In this study, for the subject→intermediate brain image registration, we tested different registration algorithms including affine alignment, nonlinear registration based on finite element method (FEM) [31], SPM nonlinear registration [77], and the highly deformable algorithm [43]. We compared their performances under the multi-template strategies in the improvement of registration accuracy.

## 4.3    EVALUATION METHODS

All 70 individual brain images (testing dataset) were warped into the MNI Colin27 space using different template selection strategies and registration algorithms. For each template strategy and registration algorithm, a mean brain image was created from the resulting warped images of all 70 testing data (subject→intermediate image(s)→Colin27 registration).

For comparison, image normalization of the testing data directly into Colin27 (subject→Colin27) using the same registration methods was also performed and a mean image was created from the 70 testing data for each registration algorithm.

The smoothness of the mean image was used as a quantitative index to evaluate the registration accuracy. Misalignment from the registration algorithm and the selection of an unrepresentative template lead to a blurred mean brain image in Colin27 space. To isolate the effect of different template selection strategies on the accuracy of image normalization, we compared the mean images calculated using the same registration method (subject → intermediate image for multi-template methods; subject → Colin27 for the single template method), but different template selection strategies. The smoothness introduced by the registration method was equivalent in these mean images, when the same registration methods were used to create these mean images. Therefore, the difference in the smoothness measurement was primarily due to the different template selection strategies.

In this study, smoothness of the mean brain image was estimated using 3dFWHM in AFNI [37], based on algorithms described in Xiong et al.[78]. In Xiong's algorithm, the spatial correlation between voxels along each axis is characterized by FWHMx, FWHMy, FWHMz (FWHM, for full width half maximum of Gaussian kernel). The averaged smoothness measurement across xyz axes is calculated by $\frac{1}{3}\sum_{i=x,y,z} FWHMi$.

The performance of the multi-template normalization (subject→s-image→r-image→Colin27) was also evaluated by the averaged percent reduction (averaged across xyz axes) in smoothness compared to single template method (Colin27).

The averaged percent reduction in smoothness(%) is calculated using the following equation: $\frac{1}{3}\sum_{i=x,y,z}(M_{FWHMi}-S_{FWHMi})\Big/S_{FWHMi}*100$ where $M_{FWHMi}$ represents the smoothness measurement of the mean image using multiple template strategies along axes $i$ $(i = x,y,z)$, and $S_{FWHMi}$ is the corresponding measurement from single template method (Colin27).

For comparison, the smoothness of the anatomical template Colin27 was also estimated and used as the inherent smoothness measurement.

## 4.4    RESULTS

### 4.4.1   Partitions in the Population

The training data (100 subjects) were divided into 9 clusters or subgroups by the complete-linkage clustering algorithm using the selected similarity threshold (mean − 0.5 * standard deviation). The distribution of the training data among the 9 subgroups is shown in Figure 4.3.
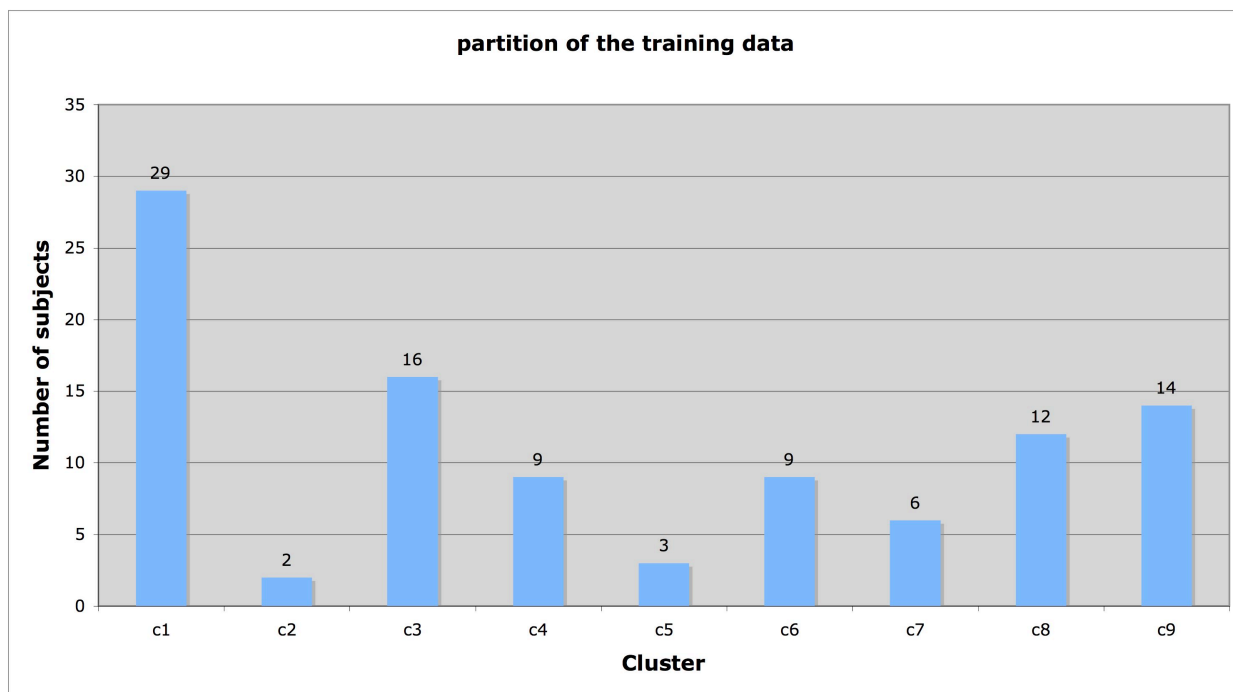
**Figure 4.3. Partitions of the training data using the complete-linkage clustering algorithm.**

All 9 members of cluster 4 are displayed in Figure 4.4, and the representative image of

cluster 4 is marked with a red square. The image of Colin27 is also shown for comparison.
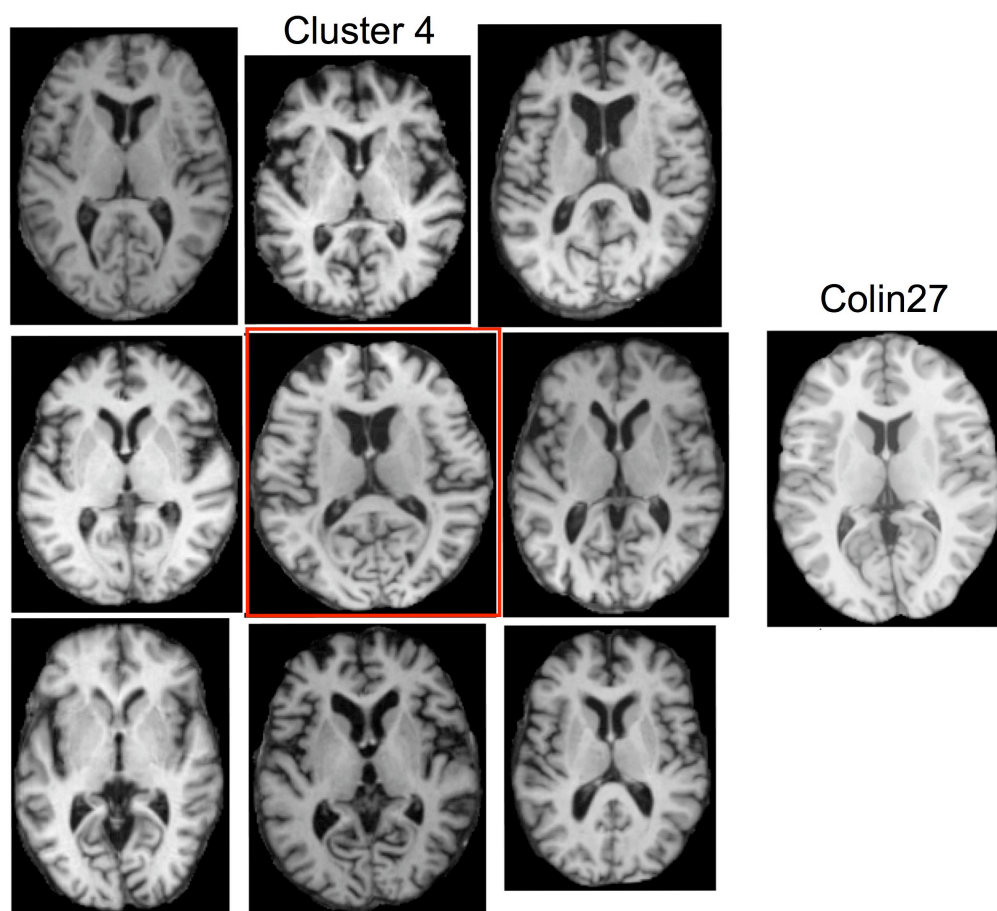
**Figure 4.4. All 9 members in cluster 4 are shown here, with the representative image outlined in red. Colin27 is also shown as a comparison.**

### 4.4.2 Performance of Template Selection Strategies

The mean images averaged from all 70 testing data using different registration methods and template selection methods are shown in Figure 4.5; Colin27 is also shown for comparison.
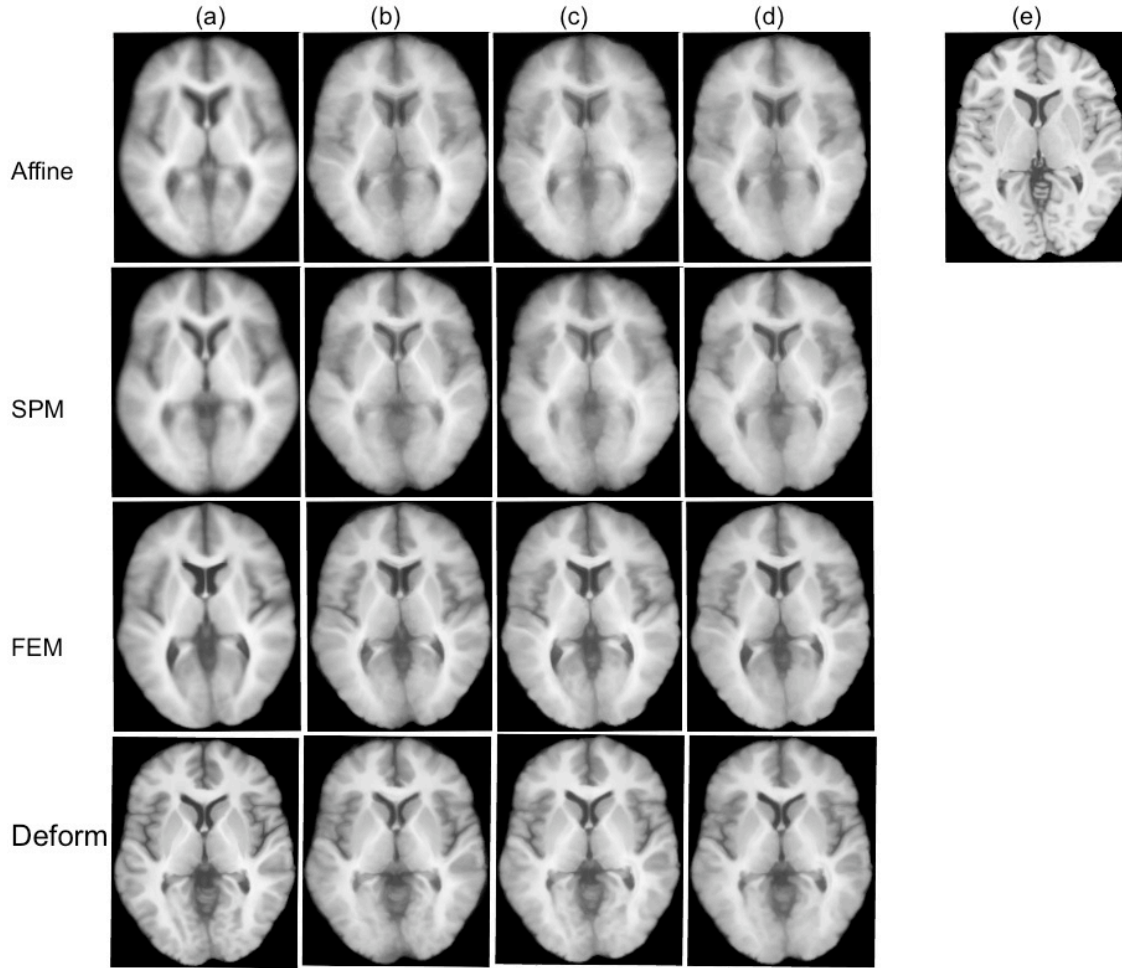
**Figure 4.5. The mean images using different template selection strategies and registration methods are compared to Colin27 (e). Respective mean images from (a) subject→Colin27, (b) subject→s-image→Colin27, (c) subject→r-image→Colin27, and (d) subject→s-image→r-image→Colin27 are shown.**

Average smoothness of the mean images in Figure 4.5 was quantitatively measured using 3dFWHM (AFNI), and the measurements are shown in Table 4.2.

The performance of the multi-template image normalization (subject→s-image→r-image→Colin27) was evaluated under different registration methods (Affine, SPM, FEM and ALP), and compared to the single template image normalization registration method using the fully deformable registration algorithm (subject→ Colin27). The smoothness measurements are

shown in Table 4.3. The inherent smoothness of the MNI Colin27 is also included in Table 4.3 for comparison.

**Table 4.2. The smoothness measurement of the mean images.**

| ▲Averaged smoothness | *Colin27 | ◆s-image | ▲r-image | ●s & r |
|---|---|---|---|---|
| Affine | 14.66 | 11.85 | 11.81 | 10.89 |
| SPM | 13.28 | 12.42 | 13.64 | 11.16 |
| FEM | 9.68 | 9.35 | 8.61 | 8.68 |
| Deform | 7.65 | 8.58 | 9.07 | 8.58 |

▲ Average smoothness of the mean image is measured as averaged full width half maximum width (FWHM) of Gaussian kernel along axis x, y or z by 3dFWHM in AFNI.

* Colin27 as the single template for the image normalization (subject→Colin27).

◆ s-image (the most similar image) as the intermediate image for the image normalization (subject→s-image→Colin27).

▲ r-image (the representative image of the cluster) as the intermediate image for the image normalization (subject→r-image→Colin27).

● s & r-image (s-image and r-image) as the middle templates for the image normalization (subject→s-image→r-image→Colin27).

**Table 4.3. The average smoothness of the mean image (multi-template vs single template)**

| Smoothness | ◆Subject→s-image→r-image→Colin27 | | | | *Deform | ▲Colin27 |
|---|---|---|---|---|---|---|
| | Affin | SPM | FEM | ALP | | |
| FWHM x | 12.63 | 9.69 | 7.78 | 7.91 | 7.15 | 5.4 |
| FWHM y | 15.95 | 12.15 | 9.21 | 9.1 | 7.97 | 5.77 |
| FWHM z | 15.4 | 11.64 | 9.04 | 8.72 | 7.84 | 5.58 |

◆ Multi-template image normalization uses both the most similar image and the prototype image (Subject→s-image→r-image→Colin27). Different registration methods including affine, SPM, FEM and ALP were tested with the template strategy.

* Deform:  Fully deformable registration with Colin27 as the template (subject→Colin27).

▲Colin27: The inherent smoothness of the template Colin27.

The detailed smoothness of the mean images along axis x, y and z for different template selection strategies using SPM nonlinear registration is shown and compared in Figure 4.6.

65
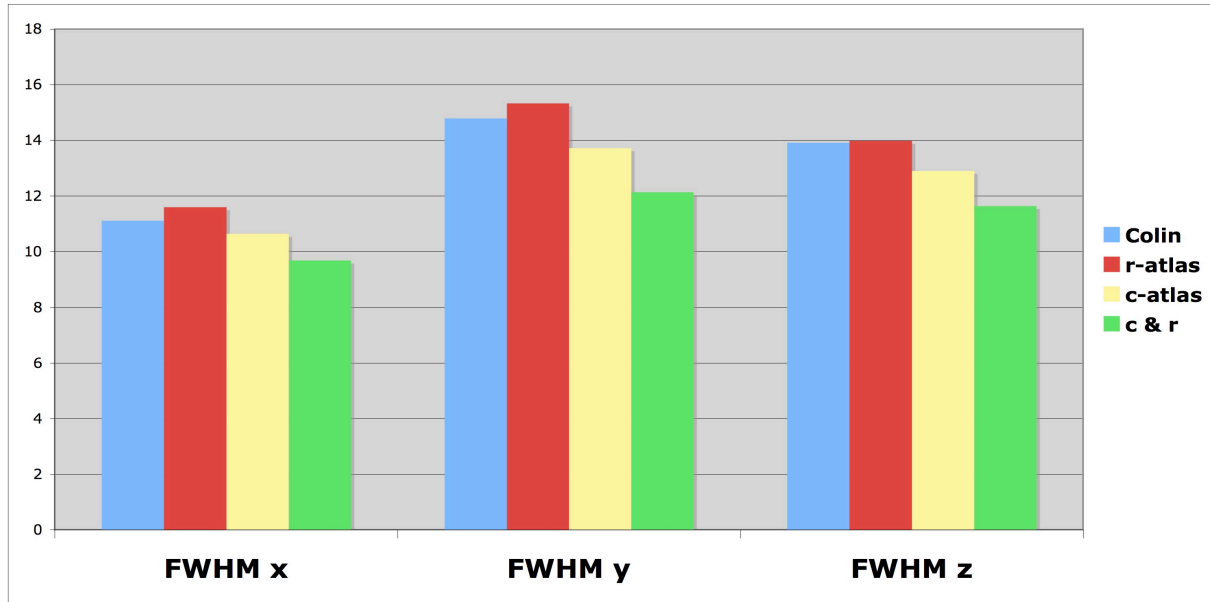
**Figure 4.6. The comparison of the performance of different atlas selection strategies under SPM.**

Average reduction in smoothness (%) —comparing multi-template strategy (subject→s-image→r-image→Colin27) to single-template image normalization under different transformation models (Affine, SPM, FEM, and a fully deformable model)— is shown in Figure 4.7.
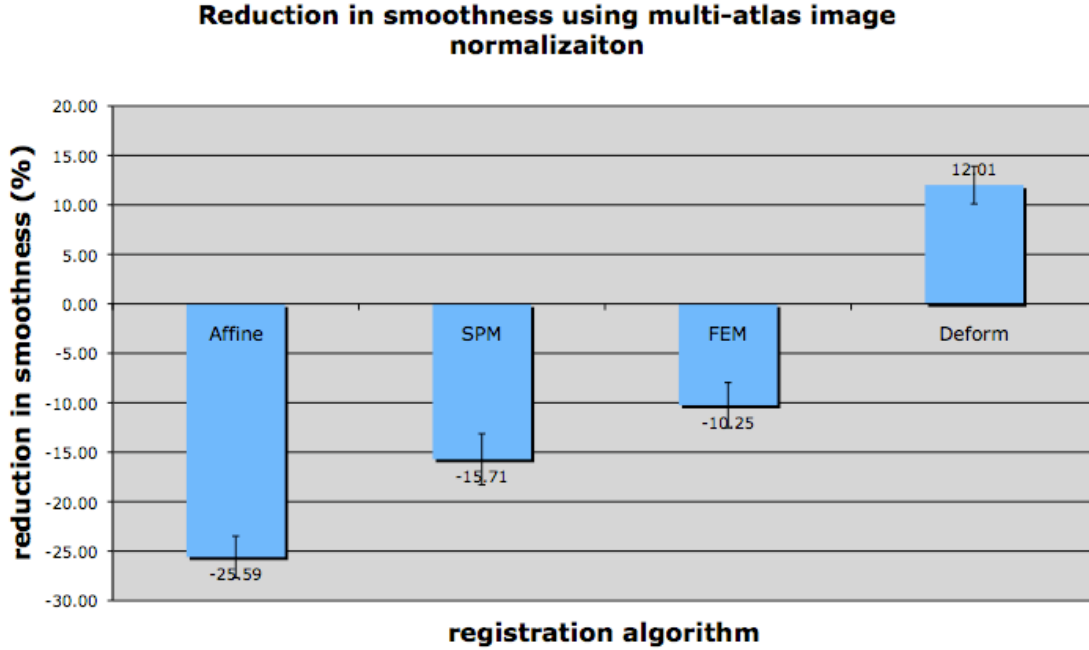
**Figure 4.7. The average reduction in the smoothness measurement (%) on the mean images (multi-template method vs single-template method).**

## 4.5    DISCUSSION

### 4.5.1    Performance of the Multi-template Strategies

The accuracy of image normalization is significantly affected by the degree of spatial transformation allowed in a given registration method. A transformation model with a higher degree of freedom leads to a more accurate alignment. This is confirmed in our results (Figure 4.5 and Table 4.2). For the same template selection strategy, the mean image becomes visually sharper (Figure 4.5) and the quantitative smoothness measurement improves (Table 4.2) as the registration algorithm changes from Affine to SPM and FEM.

67

Other than the registration method, the template selection strategy also affects the registration accuracy. For the linear and nonlinear registration algorithms (Affine, SPM nonlinear and FEM nonlinear models), the multi-template strategies consistently provided better image normalization results in both the visual and quantitative evaluations (Figure 4.5 and Table 4.2). For the tested registration algorithms (Affine, SPM and FEM), the multi-template strategy (subject→s-image→r-image→Colin27) outperformed the other template selection strategies by providing sharper mean images and better smoothness measurement.

For the same registration algorithm, it appears that variability in individual anatomical structures of the elderly population leads to a lower registration accuracy when using an unrepresentative single-brain template; introducing multiple intermediate brain images in the image normalization can reduce the misalignment caused by this variability. The ability of the multi-template strategy (subject→s-image→r-image→Colin27) to improve registration accuracy differs depending on the choice of registration algorithm, since registration algorithms have different abilities at compensating for the anatomical variations in the population. For example, affine registration only corrects the global misalignment using 9 independent parameters, which makes it vulnerable to anatomical variations. As expected, the multi-template image normalization significantly improves the registration accuracy by 25.6% (25.6% reduction in smoothness measurement of the mean image) when using affine registration, compared to 15.7% for SPM nonlinear registration and 10.3% for FEM nonlinear algorithm. However, contrary to our prediction, no improvement in image registration with the multi-template strategy was found when the highly deformable registration was used. Instead this approach led to a more blurred mean image (-12.0%), as indicated in Table 4.3 and Figure 4.7. Possible reasons for this finding are: 1) The registration of intermediate images and Colin27 (s-image→r-image, s/r-

image→Colin27) use the same highly deformable registration algorithm, and thus it does not take advantage of the multi-template strategy, and 2) When the registration algorithm progresses to a high flexibility, the algorithm may be able to compensate for most, if not all, of the anatomical variations in the population. In this case, the multi-template strategy may not help improve the registration accuracy. However, previous studies have shown that the highly deformable algorithm is not perfect (Wu, et al. 2007). The most likely reason for our finding is related to the shared use of the highly deformable registration algorithm by Colin27 and the intermediary registration approach.

The multi-template based image normalization significantly improved the registration accuracy for linear and nonlinear registration algorithm by 25.6-10.3%, and lowered the registration accuracy by 12.0% for the fully deformable model (Deform). To improve the registration accuracy and take full advantage of the multi-template strategies, other registration methods can be tested, such as landmark based image registration or FreeSurfer [79, 80]. However, these methods are either labor-intensive or exceptionally expensive in computation for the offline registration of the intermediate images→Colin27.

In this study, the multi-template strategy was evaluated in the application of image normalization, but it could easily be transferred to atlas-based segmentation. It would be of interest to evaluate the performance of the multi-template strategy in atlas-based segmentation. Especially this approach might be computationally less expensive than the multi-atlas segmentation approach based on multiple-classifier fusion or optimum template selection [51-55].

## 4.6    SUMMARY

Given the anatomical brain variations in the elderly population, template selection is an important and challenging decision, since it can significantly influence the final registration accuracy. The single-subject template strategy fails to characterize the variability of brain anatomical structures in the elderly population, and the averaged or statistical templates lack anatomical details. Therefore, in this study, several multi-template strategies have been proposed and evaluated for the application of image normalization using linear, nonlinear and deformable registration algorithms. Based on the results specified in section 4.4, the multi-template strategy (subject→s-image→r-image→Colin27) can significantly improve the registration accuracy for linear and nonlinear registration algorithms. This multi-template method better represents the anatomical variations in the elderly population, and also retains the detailed structures in the templates. Also in the multi-template strategy, the registration between the training data (s-image→r-image) and Colin27 (r-image→Colin27) can be computed offline to take advantage of computationally expensive registration algorithms.

# 5.0    SEGMENTATION AND LOCALIZATION OF WHITE MATTER HYPERITENSITIES

A number of previous studies have shown that white matter hyperintensities (WMHs), also called leuokoaraiosis and commonly seen on T2-weighted FLAIR MR images, are associated with neuropsychiatric disorders including vascular dementia [27], Alzheimer's disease [28], and late-onset late-life depression [29, 81].

This chapter will describe the implementation and validation of an automated method for quantifying and localizing WMHs. A fuzzy connected algorithm [82, 83] is adapted to automate the segmentation of WMHs and a fully deformable image registration (Chapter 2) is used to automate the anatomic localization of the WMHs using the Johns Hopkins University White Matter Atlas.

The results of a quantitative WMH assessment of a group of elderly control subjects compared to a group of Late-onset Late-life Depression (LLD) subjects are presented in this section. The LLD group was chosen because it is known that these subjects have a high WMH burden [84]. Also in this section, the WMH volumes identified with our approach are compared to the gold standard assessments based on manual expert ratings. Additionally, the anatomical localization of the WMHs found with our approach is described, and the WMH burden of the control group is region-wise statistically compared to that of the LLD patient group.

## 5.1    TRADITIONAL WMH QUANTIFYING METHODS

Two analytic strategies have been used to evaluate WMHs on MR brain images: (1) semi-quantitative rating systems and (2) quantitative volumetric analyses. For semi-quantitative systems, the WMHs are visually graded by trained expert raters. The rater assigns each MR image a WMH severity score based on its visual similarity to 'prototype' MR images. Typical rating scales range from low to high severity using 4-point or 10-point scales [85-87]. This method requires subjective judgment since it describes the WMHs using 4 or 10 crude grades. It does not provide accurate information about the location or volume of the WMHs, and thus may not detect some subtle WMH differences across groups. Also, different visual rating scales make it difficult to compare or reproduce the findings on WMHs evaluation across medical centers [88].

For quantitative analyses on WMHs, several methods have been explored to automatically or semi-automatically segment the WMHs. For example, the K-Nearest Neighbor (KNN) classification method was used to automatically or semi-automatically label the T2-weighted MR brain images as gray matter, CSF and white matter lesions [89-92]. In this method, the classification of an image voxel from a new patient relies on the voxel intensities and spatial information of a previously manually classified training set. Since the MR image of different subjects at the same medical center or across medical centers may have different intensity distribution ranges, and the normal anatomic variations across subjects lead to variability in the spatial features, this method may not work for some subjects. Other machine learning algorithms including artificial neural networks [93] have also been investigated for WMH segmentation, though they face similar dependencies of requiring a training set. An automated method has been proposed [94] that can delineate large brain lesions on T1-weighted structural images, which

involves comparing the smoothed individual T1-weighted image to a control group using the general linear model (GLM). The accuracy of this method depends on the performance of the spatial normalization technique. The normal anatomical variations in brain structures between individual subjects and the control group may present a problem for the registration accuracy and GLM. Thus, a Gaussian smoothing filter is used to smooth out the anatomical differences, which may also affect the reliability of the volumetric quantification of the lesions [94].

On T2-weighted FLAIR MR images, the WMHs usually have a higher intensity than normal white matter (WM). Some methods automatically or semi-automatically segment the WMHs on FLAIR images by defining a cut-off threshold on the images. For example, 3.5 standard deviations (SD) of the intensity value of the normal WM has been used as the lower intensity threshold for WMH segmentation [95]. The histogram of the FLAIR image has been used in a regression model to decide on a cut-off intensity threshold, and the voxels above the threshold are classified as WMHs [96]. Another method uses the mean and standard deviations of the gray matter, white matter and Cerebrospinal fluid (CSF) to estimate the intensity threshold for WMH, in which a probability map is used to favor the most likely WM regions [97]. These methods use only a single intensity threshold to segment the WMHs for the whole brain or for each slice of the brain images, which may misclassify some non-WMHs as WMHs, since some gray matter demonstrates signal intensity above the threshold [95] and the image intensity inhomogeneities may be problematic. To exclude the misclassified voxels, a manually outlined mask of WMHs with surrounding WM, GM and CSF has been used in Hirono's paper, while in Wen's paper a WM probability map (MNI 152 brains) has been used to favor the most likely WM regions. Manually outlining the WMH mask of a 3D brain volume is time-consuming and

labor-intensive, while using a WM probability map in a MNI template will make the accuracy of WMH segmentation dependent on the quality of the inter-subject registration.

Previous research suggests that the location or distribution of WMHs is associated with specific symptoms [98]. Most previous studies have focused only on WMH visual inspection or volume measurement and did not distinguish among anatomically distinct WMHs, while a few groups have explored semi-automated or automated methods to localize WMHs into large compartments or categories such as periventricular white matter hyperintensities (PVWMHs) and deep white matter hyperintensities (DWMHs). For example, in Swatz et al. [91], a 3D classification algorithm was applied to separate DWMHs from PVWMHs. Other investigators have used nonlinear image registration methods to convert the WMHs across subjects into a standard space [99, 100].

## 5.2    AUTOMATED WMH SEGMENTATION AND LOCALIZATION

This section describes our test data set and methods for the WMH segmentation and localization procedure. The data set description includes descriptions of (1) subjects, (2) MR Imaging Parameters, and (3) ) 'gold standard' White Matter Hyperintensity Ratings. The major steps of the methods involve (1) image preprocessing, (2) automated WMH segmentation, and (3) automated WMH localization.

### 5.2.1 Test Data Set

### 5.2.1.1 Subjects

Nineteen subjects (eleven patients and eight controls) were recruited through the University of Pittsburgh Intervention Research Center for Late-Life Mood Disorders. Subjects were 63 to 81 years of age (mean age= 72.3, S.D.=4.86), and their WMH visual scores ranged from 0.5 to 6.5 (mean WMH score=2, S.D.=1.6). All subjects (controls and depressed) received a SCID-IV evaluation, which was reviewed in a diagnostic consensus conference. Eleven of the 19 subjects were diagnosed as depressed patients; while the remaining eight subjects were termed control subjects. The 11 patients had late-onset late-life depression; they met DSM-IV criteria for Major Depressive Disorder [101], and their depression began at the age of 60 years or older. The mean Hamilton Depression Rating Scale on patients was 20.3 (S.D.=4.9). The subjects did not have significant cognitive impairment and their mean Mattis Dementia Rating Scale was 136.3 (S.D.=5.9). They were all participants in a research trial of antidepressant medications. Other than Major Depressive Disorder (for subjects in the depressed group) and anxiety disorders, all other Axis I psychiatric disorders were used as exclusion criteria. We chose to include subjects with co-morbid anxiety disorders due to the high prevalence (48%) of anxiety disorders in subjects with late-life depression [102]. Each subject was assessed by the Mini Mental State Examination (MMSE)), Hamilton Rating Scale for Depression (Hamilton), and Mattis Dementia Rating Scale (Mattis). Clinical characteristics of the subjects (patients and controls) are summarized in Table 5.1. The 2 groups were well balanced with respect to gender and age. The MR images used in the current analyses were obtained at the time of subject's enrollment, before the antidepressant medication was started. This study was approved by the University of Pittsburgh Institutional Review Board (IRB). Written informed consent was obtained.

**Table 5.1. Clinical characteristics of the subjects.**

|  | Group I (Depressed) | Group II (Controls) | *t*-test probability |
|---|---|---|---|
| No. of subjects | 11 | 8 | |
| Age, year (range) | 72.2±5.3 (63-80) | 72.3±4.8 (67-81) | 0.936 |
| Gender, M/F | 5/6 | 4/4 | |
| WMH scores ± SD | 2.55±1.9 | 1.25±0.5 | 0.054 |
| MMSE ± SD | 27.7±3.6 | 28.8±1.5 | 0.406 |
| Hamilton ± SD | 20.3±4.9 | 2±2.07 | 2.248E-08 |
| MATTIS ± SD | 136.3±5.9 | 139.9±3.4 | 0.116 |

Statistical comparisons utilized a two-sample, unequal variance, two-tailed Student's *t*-test.

## 5.2.1.2 MR Imaging Parameters

Magnetic resonance images were acquired on a 1.5 Tesla Signa Scanner (GE Medical Systems, Milwaukee, WI). The 3D structural MR images were acquired at sagittal orientation using 3D Spoiled GRASS (SPGR, TR/TE = 5/25 ms; flip angle = 40º; FOV = 24×18cm, slice thickness = 1.5mm, matrix = 256×192 matrix).

The following axial series were also obtained: T1-weighted (TR/TE = 500/11 ms, Nex = 1) and fast fluid-attenuated inversion recovery (fast FLAIR) (TR/TE 9002/56 ms Ef; TI = 2200 ms, NEX = 1). Section thickness was 5 mm with a 1-mm inter-section gap. All axial sequences were obtained with a 24 cm field of view and a 192 x 256 pixel matrix. Slice thickness and orientation were chosen so that the acquired images would be compatible with the WMH rating scales described below.

### 5.2.1.3 White Matter Hyperintensity Ratings

The 'gold standard' WMH ratings are based on a system developed for the Cardiovascular Health Study [CHS; 85, 86]. A numerical rating for the WMHs was assigned based on the comparison of each subject's imaging data to predefined CHS visual standard, representative of progressive severity within a 10-point scale (0 through 9). In this study, two raters independently evaluated WMH on the FLAIR images. If they differed in their ratings by one point, the final rating was the mean of the two values. A greater than one-point difference between raters was considered as a disagreement, and was adjudicated by consensus.

### 5.2.2   Methods

### 5.2.2.1 Image Preprocessing

Image preprocessing includes skull stripping of the SPGR and FLAIR brain images, which improves the accuracies of WMH segmentation and localization. For the skull stripping on the FLAIR images, the Brain Extraction Tool [BET, 103] was used on the T1-weighted images, which were acquired at the same location and voxel-size as the FLAIR images. The resulting stripped T1-weighted image was then used as a brain mask to remove the skull and scalp from the FLAIR image.

### 5.2.2.2 WMH Segmentation

The automated WMH segmentation method involves four steps: (1) automatically identifying WMH seeds based on the intensity histogram of the FLAIR image, (2) using a fuzzy connected algorithm to segment the WMH clusters, (3) iteratively updating the set of seeds, and (4) combining the WMH clusters into the final WMH segmentation. The histogram of the skull-

stripped FLAIR image is used to define a threshold (mean + 3 × SDs) for seed selection; voxels

beyond this threshold are classified as WMHs and are used as seeds in the fuzzy connected

algorithm to segment surrounding WMH voxels. The background of the FLAIR image is

excluded when calculating its intensity histogram, mean intensity and standard deviation. In the

fuzzy connected algorithm, the fuzzy adjacency and affinity, both between 0 and 1, are defined

for each pair of voxels $(a,b)$: the fuzzy adjacency $\mu_\alpha(a,b)$ defines how close the two voxels are,

while the affinity $\mu_k(a,b)$ (determined based on adjacency degree $\mu_\alpha(a,b)$ and intensity

similarity) indicates how strongly the two voxels "hang together" in space and intensity. A fuzzy

connected object is a set of voxels $O$ with properties as follows: any two voxels $(a,b)$ from $O$

have an affinity $\mu_k(a,b) > x$ where $0 \leq x \leq 1$, and for any pair $a \in O, b \notin O$, the affinity is

$\mu_k(a,b) < x, 0 \leq x \leq 1$. A detailed and precise mathematics definition is given in Udupa, et al.

[104] and Udupa, et al. [105]. For each selected WMH seed the fuzzy connected algorithm

generates a fuzzy object, within which each pair of voxels has a strong fuzzy connectedness or

affinity (above certain threshold, 0.5 in this study), and the system automatically delineates a 3D

WMH cluster containing the respective seed. Multiple 3D FLAIR image WMH clusters are

generated from the set of automatically selected seeds and then combined to form an overall

WMH segmentation volume.

The flow chart of the WMH segmentation is shown in Figure 5.1. The fully automated

WMH segmentation system was implemented in C++ and ITK. The WMH segmentation

algorithm       is       available       upon       request       through       our       website

(http://www.pitt.edu/~aizen/GPN_Home.html).
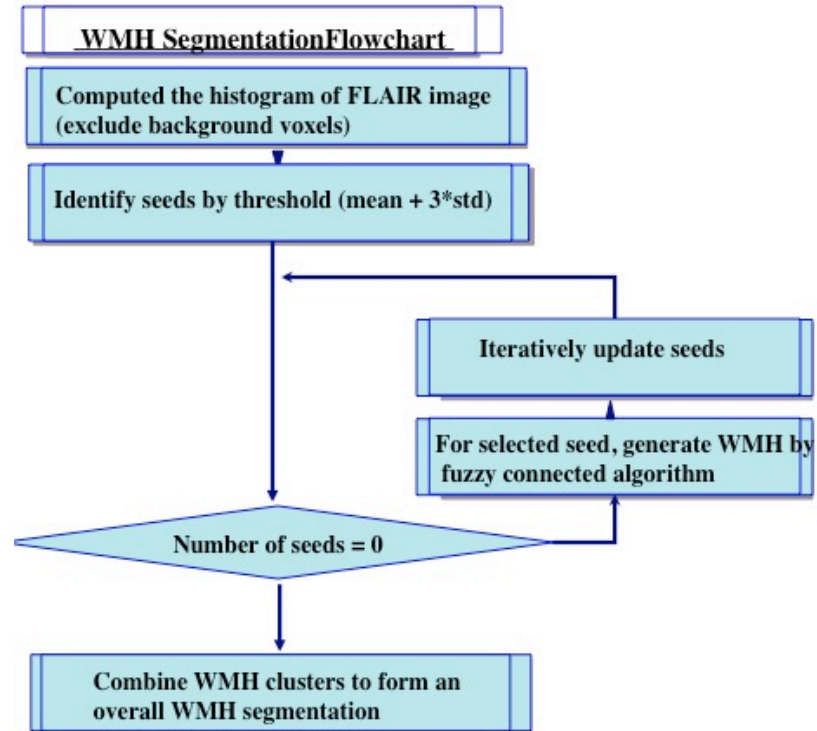
**Figure 5.1. WMH Segmentation Flowchart. The processing steps used to automatically segment the WMHs on FLAIR MR brain images.**

### 5.2.2.3 WMH Localization

Automated Labeling Pathway (ALP, see Figure 5.2) is an automated method developed over a series of functional and structural MRI studies to automatically label specific anatomic regions of interest [106-108]. The pathway combines a series of publicly available software packages such as AFNI (Cox, 1996), BET (Smith, 2002), FLIRT (Jenkinson, 2002) and ITK (Yoo, 2004), as well as some locally developed programs to automatically label ROIs on the SPGR image of a subject.
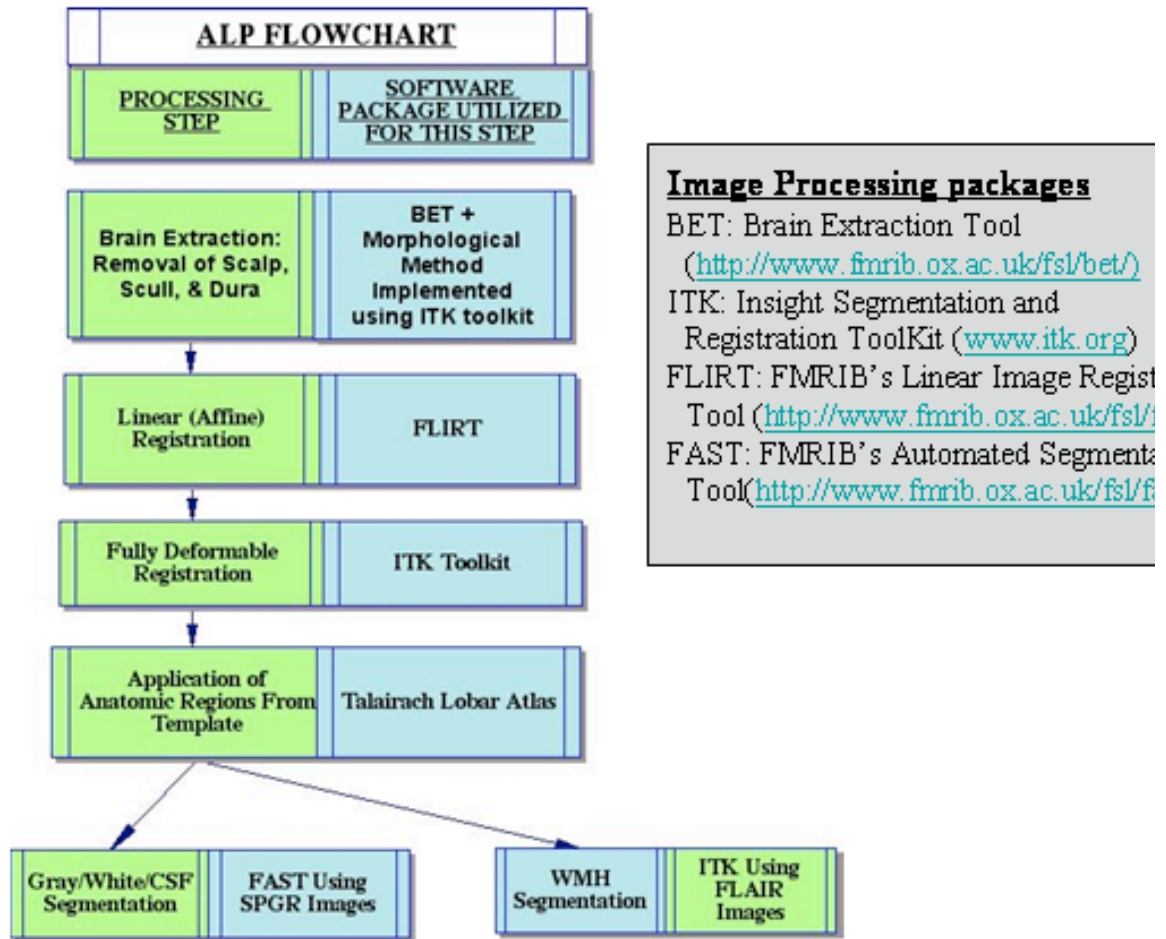
**Figure 5.2. ALP Flowchart. The processing steps that constitute the Automated Labeling Pathway (ALP), which is used to generate regional brain volume estimates. The process uses a variety of publicly available packages, as well as some locally developed programs, for atlas-based segmentation of MR images.**

The deformable model described in Chapter 2.0 is used in ALP for inter-subject registration (template colin27→subject 3D SPGR). The deformable registration allows for a high degree of spatial deformation, which seems to give it a particular advantage over other standard registration packages, such as Automated Image Registration (AIR) and Statistical Parametric Mapping (SPM).

An overview of the WMH localization procedure is summarized in Figure 5.3. The high-resolution reference image (MNI colin27) is registered to the T1-weighted SPGR high-resolution image of the subject using ALP, and the Johns Hopkins University White Matter Atlas (defined

on the reference brain MNI colin27 image) is warped into each individual's anatomic image space. The anatomic information in subject SPGR space is then transformed further into the subject's FLAIR image space by rigid-body registration between the subject SPGR image and subject T1 in-plane image, which was acquired using the same slice prescription as the subject's FLAIR image. In this way, the anatomical information in the atlas is carried into the subject's FLAIR space and the ROIs labeled on the subject's FLAIR image are used as binary masks to localize the WMHs. The WMH localization task is viewed as a registration procedure. The Johns Hopkins University White Matter Atlas used in the current study is based on high-resolution diffusion tensor MR imaging and 3d tract reconstruction. The atlas has 21 prominent white tracts including anterior thalamic radiation (ATR), cingulum (Cg) and other tracts [109], as listed in Table 5.2.
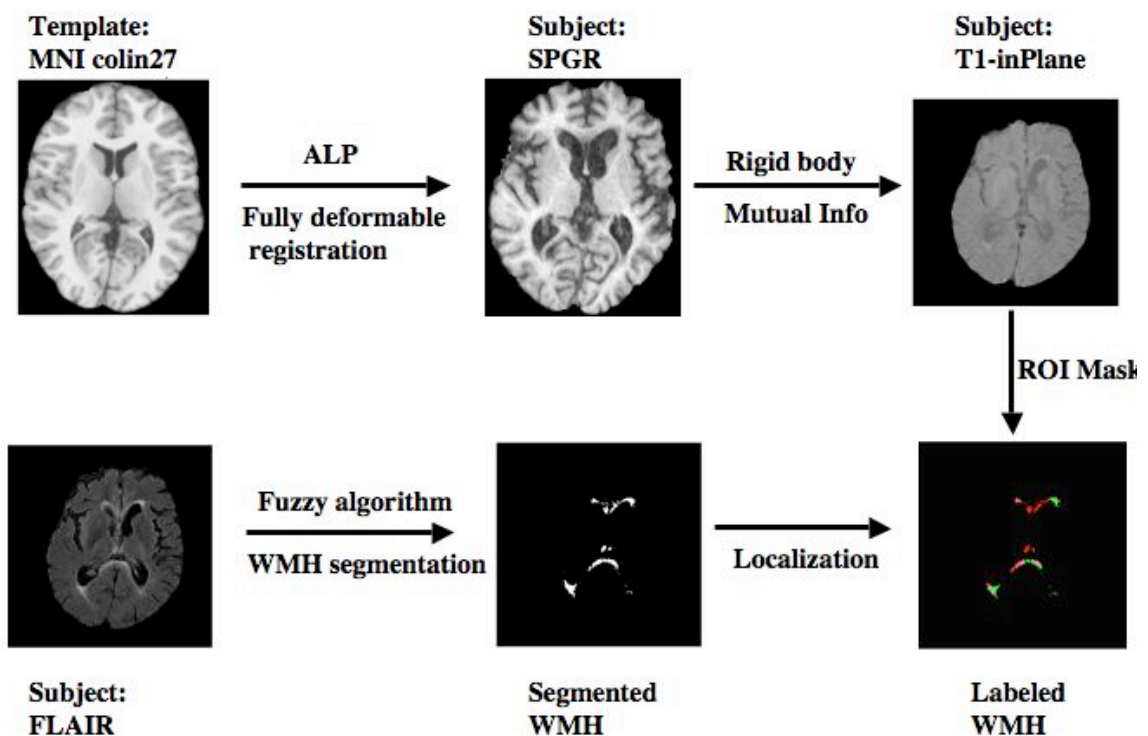


**Figure 5.3. An overview of the WMH localization procedure.**

## 5.3    RESULTS AND DISCUSSION

### 5.3.1   WMH Segmentation Evaluation

A subject with some discrete lesions as well as confluent lesions was chosen to demonstrate the results of this WMH extraction algorithm. Nine pairs of the segmented WMH slices versus corresponding FLAIR slices from the subject are displayed in Figure 5.4, showing this method's effectiveness in the segmentation of discrete as well as confluent WMHs.

The WMH segmentation results of 19 subjects using this automated method were statistically compared to the WMH visual grades from the manual ratings. The comparison was done with a linear regression model.  In this study we chose to use semi-quantitative CHS ratings as the gold standard for comparison.  An alternative approach would have been to use manually segmented WMH tracings.  Since the two measures being compared used different metrics, we are only demonstrating a correlation between the measures rather an absolute agreement.

The WMH volumes of the 19 subjects from the automated segmentation method were found to be significantly correlated to the visual grades with a $R$-squared $= 0.909$ and $F(1,18) = 170.7$, $P < 0.0001$. Since the visual grade is a global index to the WMH severity on the subject brain image, the WMH volume is normalized by the overall WM volume (calculated from SPGR brain image). The normalized WMH results were also significantly correlated to the visual grades ($R$-squared $= 0.909$, $F(1,18) = 170.3$, $P < 0.0001$). This WM normalization method may not be the best way for whole brain adjustment, since previous studies have showed that WMH are significantly related to atrophy [110, 111]. A whole brain normalization method, which takes brain atrophy into consideration, may be better for WMH assessment.

The high correlation between the normalized WMH quantifications from the automated method and the visual grades demonstrates that this automated method can successfully segment the WMHs on MR FLAIR images.
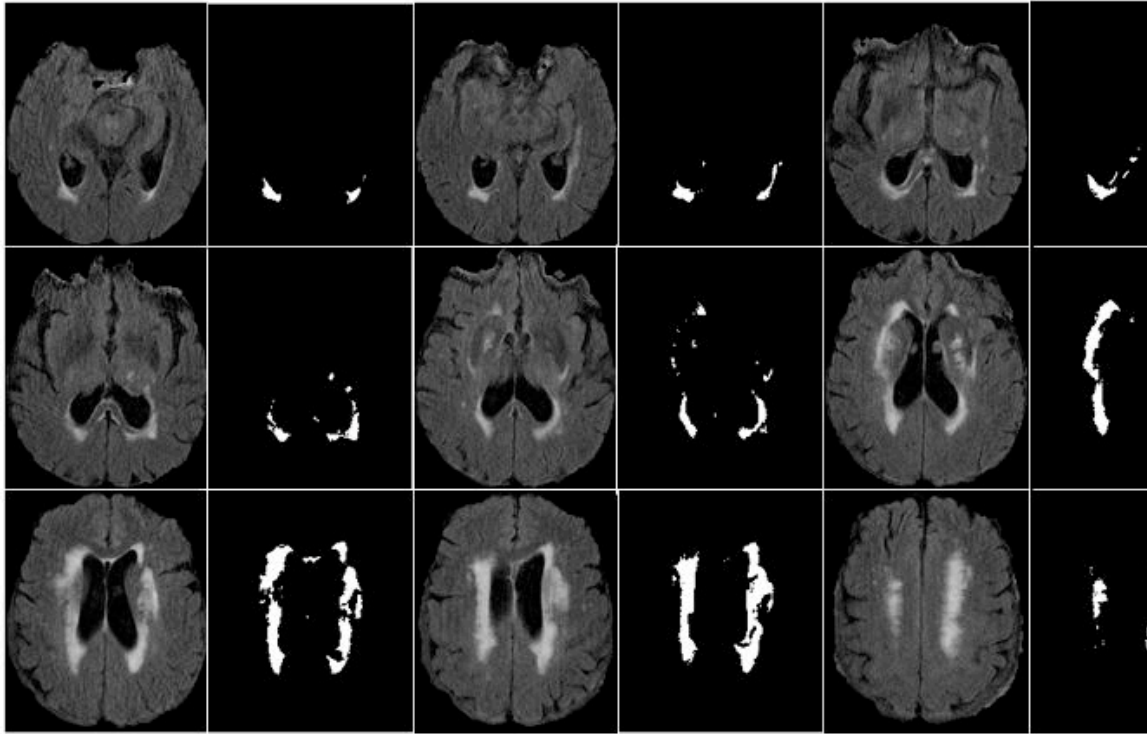


**Figure 5.4. Automated WMH segmentation results on the FLAIR MR images of one subject. Nine paired image slices on the subject are shown here. In each paired slices, the left slice is the FLAIR slice and the right one is the associated automated WMH segmentation result.**

### 5.3.2   Localization of WMHs

Using ALP, the Johns Hopkins University White Matter Atlas is transferred to each subject's 3D SPGR image and further carried into subject FLAIR image space. The atlas regions in the subjects' FLAIR image space are then used as ROI masks to localize the WMHs. Figure 5.5

shows the segmented ROIs in MNI template colin27 space, individual SPGR structural space and

FLAIR image spaces; respective MR images are also shown as underlay images.



(a)  (b)  (c)
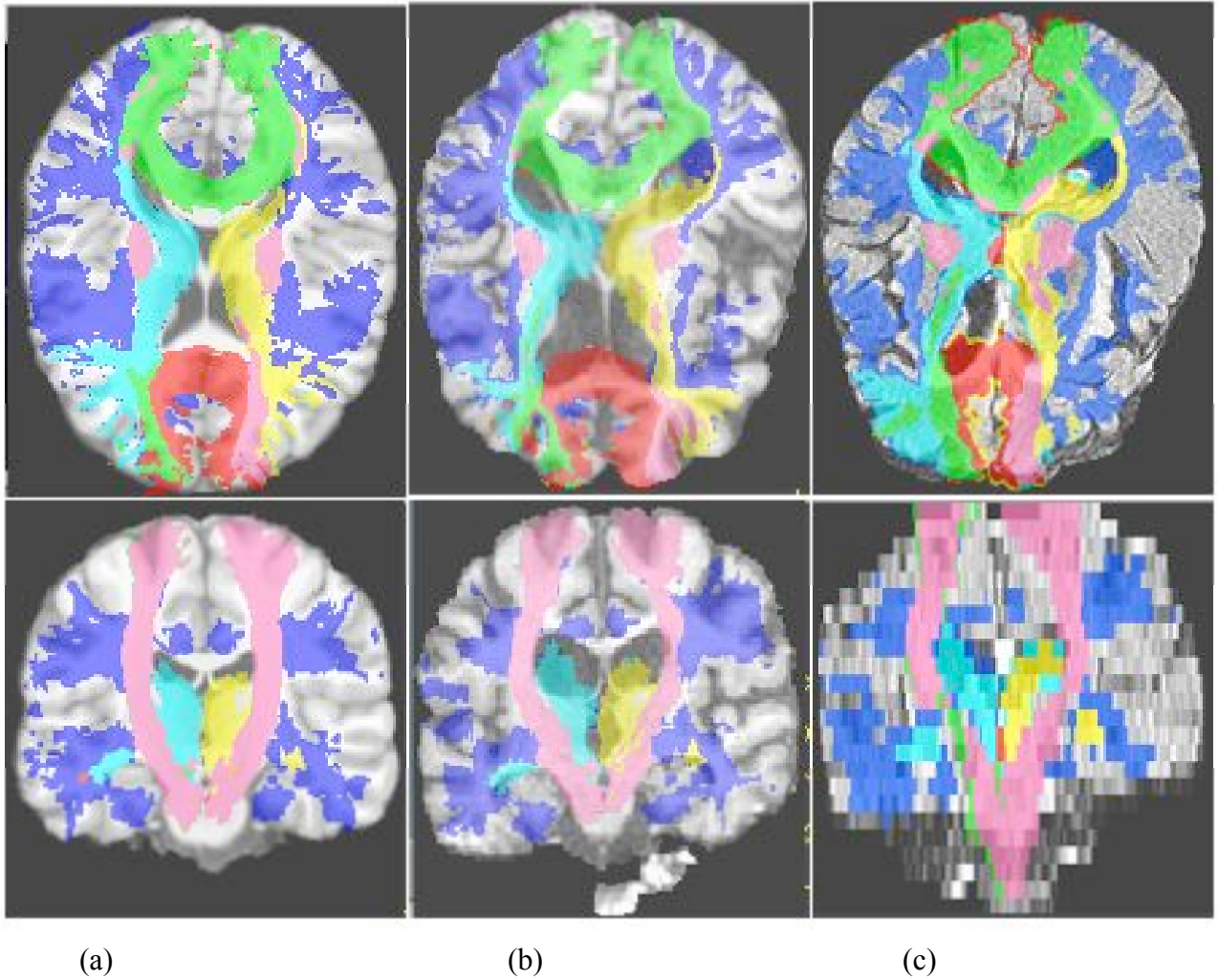
**Figure 5.5. The result of atlas-based segmentation from ALP. Segmentation results are shown at axial orientation in the top row and the coronal orientation in the bottom row. (a) The MNI template colin 27, overlapped with the Johns Hopkins University White Matter Atlas (i.e., Anterior thalamic radiation, corpus callosum, corticospinal tract, inferior fronto occipital, inferior longitudinal fasciculus, superior longitudinal fasciculus, right uncinate fasciculus etc. (b) A single subject 3d SPGR image, overlapped with the transformed ROIs. (c) The same single subject FLAIR image, overlapped with the transformed ROIs.**

The localized WMH volumes are quantified by multiplying voxel size by the number of

WMH voxels inside the ROIs including anterior thalamic radiations, corticospinal tracts etc, as

listed in Table 5.2. The WMH volume estimates from WMH localization describe the spatial

distribution of the WMH burden, which can facilitate further research on the role of WMH in pathogenesis of neuropsychiatric disorders. In Table 5.2, for each region of interest, the WMH volumes of the LLD patient group were statistically compared to the WMH volumes of the control group using two-tailed two-sample unequal variance $t$-test. A significant difference was found in whole brain WMH volume between the LLD patient group and the control group. The results from the WMH localization method provide more anatomical specificity. As shown in Table 5.2, there were significant differences in WMH spatial distributions between the LLD patient group and the control group in regions including right anterior thalamic radiation, corpus callusum (CC), inferior fronto occipital (IFO), inferior longitudinal fasciculus (ILF), and right uncinate fasciculus (UNC). However no significant differences were found in cingulum (CgLL, CgLR, CgUL, CgUR) and superior longitudinal fasciculus (SLFBL, SLFBR, SLFTL, SLFTR).

**Table 5.2. Mean Volumes of WMH (mm$^3$) per region for the control group (8 controls) versus the patient group (11 patients) and the $t$-test results on normalized WMHs.**

| Region\ Group | Mean WMH Volume Control (mm$^3$) | Mean WMH Volume Patient (mm$^3$) | Two tailed $t$-test on Normalized WMH ($p$ value) |
|---|---|---|---|
| Whole Brain | 2737.3 | 8541.0 | 0.0427 |
| ATRL | 649.8 | 2211.4 | 0.0554 |
| ATRR | 821.2 | 1764.8 | 0.0187 |
| CCF | 201.7 | 1165.6 | 0.01687 |
| CCO | 416.1 | 1529.6 | 0.01277 |
| CSTL | 148.3 | 309.2 | 0.05797 |
| CSTR | 158.3 | 652.5 | 0.0609 |
| CgLL | 35.1 | 110.3 | 0.0811 |
| CgLR | 8.7 | 40.4 | 0.0606 |
| CgUL | 11.9 | 550.1 | 0.1364 |
| CgUR | 24.2 | 29.5 | 0.8187 |

Table 5.3. (continued)

| | | | |
|---|---|---|---|
| IFOL | 338.1 | 1226.8 | 0.03430 |
| IFOR | 478.7 | 2039.0 | 0.0391 |
| ILFL | 192.6 | 716.6 | 0.0186 |
| ILFR | 214.9 | 860.4 | 0.0384 |
| SLFBL | 273.3 | 929.7 | 0.3593 |
| SLFBR | 157.0 | 1441.2 | 0.3005 |
| SLFTL | 128.7 | 717.5 | 0.4059 |
| SLFTR | 46.5 | 761.5 | 0.2977 |
| UNCL | 104.0 | 430.3 | 0.1685 |
| UNCR | 139.2 | 327.8 | 0.0304 |

Keys: ATRL/R-anterior thalamic radiation (left or right), CCF/O-corpus callosum (frontal or occipital), CSTL/R-corticospinal tract (left or right), CgLL/R-cingulate (lower part left or right), CgUL/R-cingulate (upper part left or right), ILFL/R-inferior longitudinal fasciculus (left or right), IFOL/R-interior frontooccipital fasciculus (left or right), SLFBL/F- entire superiro longitudinal fasciculus (left or right), SLFTL/F-superiro longitudinal fasciculus (the branch to the temporal lobe, left or right), UNC-uncinate fasiculus(left or right).

### 5.3.3 Limitations and Advantages

The current study is limited by the low-resolution FLAIR image, as well as the limited number of subjects (11 patients and 8 control subjects). The analyzed FLAIR images were acquired with a slice thickness of 5mm and a 1mm gap, which may be an inadequate resolution for accurate volumetric quantification of the WMHs, and accordingly may affect the reliability of the group comparison results. A higher image resolution, such as a slice thickness of 2mm with no gap, could improve the WMH quantification and improve the registration accuracy, which would lead to more accurate WMH localization. Also, a larger group of well-characterized LLD subjects with a matched elderly non-depressed control group would add confidence to the WMH localization findings in Table 2 that are specific for LLD.

Previous research has used a fuzzy connected algorithm for semi-automated WMH segmentation [105, 112], which required some user interaction and did not give spatial information on the WMHs. Our method automatically identifies WMH seeds, and generates and localizes WMH segmentation; our method is objective and does not require any manual interaction. For the segmentation, we use the histogram of the FLAIR image to automatically generate the WMH seeds, and then use the fuzzy connected algorithm with specific parameters to form a WMH cluster for each respective seed. This method updates the seeds iteratively and combines the scattered WMH clusters into the final WMH segmentation. Since the fuzzy connected algorithm uses different parameters for each seed, this method enables a different threshold for each WMH cluster and avoids a single cut-off threshold for the whole brain or brain slice. This potentially offers more precise WMH segmentation. A fully deformable registration (ALP, Chapter 2.0 and [43]), which combines the piecewise linear registration for coarse alignment with Demons algorithm for voxel-level refinement, is used for accurate WMH localization on the Johns Hopkins University White Matter Atlas [109].

The WMH segmentation and localization method we described provides more specific and more accurate information about WMH volume and spatial distribution than visual WMH grades. Unlike different visual grading systems, it is very easy to compare the WMH findings from this method across different centers. The method relies on the properties of the subject's own FLAIR image such as the intensity distribution of WMHs, and the connectivity and the diffusivity of the WMHs for the WMH segmentation, which does not rely on any training dataset like some of the reviewed methods [89-92].

## 5.4    SUMMARY

In this chapter we presented and validated a new method for fully automated segmentation and localization of WMHs on MR images. The method adapts the fuzzy connected algorithm for WMH segmentation and uses a demons-based fully deformable registration for WMH localization. The automated WMH segmentation method was evaluated by comparing the resulting WMH quantifications (non-normalized or normalized by total WM volumes) of the 19 elderly subjects (11 late-life depressed subjects and 8 elderly controls) with the standard visual grading approach for estimating WMH burden. In the comparisons a high correlation of the WMH ratings was found between our new semi-automated approach and the manual ratings. Specifically, the two methods correlate with $R$-squared = 0.909, $P$ <0.0001. Further localization of WMH followed the expected patterns of LLD: high WMH burden in the subcortical, and frontal regions.

Quantification and localization of WMH volumes is critical for research involving the risk factors and pathogenesis of neuropsychiatric disorders. Most previous methods were labor intensive, subjective, and provided little if any anatomic localization. The current method solves many of the previous limitations: it does not require any manual intervention, provides WMH volume estimates, and localizes the WMH burden to a number of anatomic ROIs. The methods described in this chapter are particularly relevant given the emergence of large MRI databases, such as that provided by the Alzheimer's Disease Neuroimaging Initiative (http://www.loni.ucla.edu/ ADNI/).

The development and implementation of an automated method for quantifying and localizing WMH will facilitate a further, fine-grained understanding of: 1) short- AND long-term

treatment response; 2) evolution of cognitive functioning in late life depression; 3) evolution of

leukoarisosis in LLD; 4) impact of medical and psychiatric treatment on WMHs in LLD; and 5)

modeling of cognitive impairment in LLD: e.g., is diminution in speed of information

processing driven primarily by WMH, beta amyloid deposition, or both?

# 6.0 RESTING STATE CONNECTIVITY AND WHITE MATTER BURDEN IN LATE LIFE DEPRESSION

In this chapter, we use the registration and segmentation methods developed in the previous chapters to study the resting state connectivity and white matter burden in late-life depression. The brain's default mode network (DMN), which activates during resting state, has been the focus of intense research. In this chapter, we study DMN activity in late-life depression (LLD) and the correlation of DMN activity changes with the White Matter Hyperintensity (WMH) burden. We hypothesize that LLD subjects will have altered DMN activity, which will correlate with the increased WMH burden.

## 6.1 PREVIOUS FINDINGS ON LATE-LIFE DEPRESSION

Late-life major depression (LLD) is associated with emotional suffering, disability, caregiver strain, suicide, and poor compliance with other medical treatments [113]. The current understanding of LLD's neurobiology is based on findings from both structural and functional neuroimaging studies. Structural changes include both reduced gray matter volumes as well as increased white matter hyperintensity (WMH) burden, which are the anatomical correlates of both neurodegeneration and cerebrovascular disease [114]. Several studies have reported significant bilateral volume reductions of the anterior cingulate, frontal cortex, hippocampus and

striatum [114-118]. Increased WMH burden in LLD has also been reported in several studies [114, 119-121]. Some [121], but not all [122] studies found that increased WMH burden is associated with poor antidepressant treatment response. The association of WMH burden with LLD led to the vascular depression hypothesis, which posits that a single vascular lesion or an accumulation of lesions may disrupt prefrontal systems that mediate both mood and executive function [123, 124].

PET functional neuroimaging studies have described changes of resting-state cerebral activity in LLD subjects when compared with non-depressed control subjects [125], however, these findings have not been consistently replicated [126]. Recently, fMRI has been adapted to examine the connectivity of the default-mode network (DMN), an organized functional network of several brain regions that are active during resting state and inhibited during the performance of active tasks [127]. The DMN is thought to reflect the brain's intrinsic ability to organize and stabilize different cerebral ensembles [128]. Analysis of resting state activity may enhance the understanding of the biological underpinning of mental illnesses pathophysiology. For example, activity in the DMN is affected in Alzheimer's disease (AD) [129], Major Depressive Disorder (MDD) [130, 131] and anxiety disorders [132]. In midlife MDD, resting state functional connectivity was significantly increased in depressed subjects compared to healthy control subjects, especially in the sub-network comprising the subgenual cingulate and the thalamus [130]. Also, the length of incident depressive episode correlated positively with the increased functional connectivity in the subgenual cingulate [130]. In contrast to the increased default-mode activity observed in midlife depression, decreased functional connectivity in the default network has been described in AD, suggesting that the pathological changes associated with AD

might reduce the DMN activity [129]. To our knowledge, there are no published studies characterizing DMN activity in LLD.

Compared with mid-life major depression, LLD is characterized by a marked heterogeneity in both phenotype (e.g., greater cognitive impairment, greater anxiety) and pathogenesis (e.g., vascular changes in the brain, neurodegeneration, monoamine dysregulation) [133]. Functional MRI with cognitive tasks has shown decreased functional connectivity in LLD, thought to reflect vascular and neurodegenerative changes [134]. Given this multilevel heterogeneity, characterizing DMN activity in LLD could help in delineating the functional neuroanatomy and pathophysiology of the disorder.

We hypothesize that subjects with LLD will show altered connectivity in the default-mode network when compared with elderly non-depressed comparison subjects. Moreover, given the prominent role of WMH in the LLD vascular hypothesis, we hypothesize that the increased burden of WMH will be correlated with the altered activity in the default network.

## 6.2    RESTING STATE CONNECTIVITY AND WHITE MATTER BURDEN IN LATE LIFE DEPRESSION

### 6.2.1   Materials

#### 6.2.1.1 Subjects

The data was collected from participants in the second Maintenance Therapies for Late-Life Depression study (MTLD-II), conducted at the University of Pittsburgh Advanced Center for Intervention and Services Research for Late-Life Mood Disorders between 1999 and 2004.

Details of the MTLD-II study protocol are described in Reynolds et al.[135]. In brief, participants had a mean age of 70.5, and were diagnosed via Structured Clinical Interview for DSM-IV (SCID) [136] with current non-psychotic, non-bipolar major depressive disorder (single-episode or recurrent), a 17-item Hamilton Depression Rating Scale (HDRS)[137] of 15 or higher, and a Mini Mental State Examination (MMSE) [138] score of 17 or higher. Subjects included in this study had a MMSE of 24 or above. Cognitive function was assessed with the Dementia Rating Scale [139]. Subjects with a clinical diagnosis of Dementia were excluded from the study.

During the acute phase of treatment in MTLD-II, subjects were treated openly with Paroxetine doses adjusted between 10 and 40 mg/day based on tolerability and response (mean (SD) final dose: 26 (11) mg/day), combined with weekly interpersonal psychotherapy (IPT). Subjects who responded to acute treatment entered a 16-week continuation phase to stabilize their response; they received the same pharmacotherapy and IPT every two weeks [135]. Response was defined categorically as a HDRS score of 10 or less [135].

Twenty-four participants, who consented to the neuroimaging protocol, had intact imaging data and a MMSE of 24 or above were included in this study: 12 elderly depressed subjects and 12 elderly comparison subjects. Four out of 12 comparison subjects and 5 out of 12 depressed subjects received antihypertensive medication. The comparison subjects did not receive psychotropic medications and the elderly depressed subjects were psychotropic-free at baseline.

**6.2.1.2 MR Imaging Parameters**

The baseline MR images were obtained at the time of subject enrollment, before initiation of pharmacotherapy. Follow-up scans were obtained after 12 weeks of pharmacotherapy, for 8 depressed subjects who responded to treatment. Post acute-treatment scans were obtained while subjects continued on a maintenance dose of Paroxetine.

There are several methods described in the literature for acquiring fMRI data for resting state. Some studies do not employ any explicit task and scan subject at rest, while others use the fMRI data from the fixation periods of block-design tasks [140], or from simple sensory-motor tasks (e.g. finger tapping). The finger tapping task requires a finger tap every 12 seconds and the brain functional activities associated with this simple task do not interfere with the DMN activity and has been used in the literature for acquiring resting-state data [129, 141]. Thus, finger tapping fMRI data was used to examine the resting state connectivity in LLD in this study.

Imaging data were collected with a 1.5-T Signa scanner (GE Medical Systems). 3D high-resolution anatomical images (SPGRs) were acquired in the sagittal orientation using 3D Spoiled Grass (SPGR, TR/TE = 5/25 ms, flip angle = 40º; FOV = 24×24 cm, and slice thickness = 1.5 mm, matrix = 256×256 voxels). Fast fluid-attenuated inversion recovery (fast FLAIR) images (TR/TE = 9002/56 ms Ef; TI = 2200 ms, FOV = 24×24cm, NEX = 1, slice thickness = 5 mm, gap = 1 mm) were also acquired for WMH volume quantification.

T1-weighted anatomical images (TR/TE = 500/11 ms, FOV = 24×24cm, slice thickness = 3.8mm, matrix = 256×256) were acquired parallel to the plane connecting the anterior and posterior commissures. Thirty-six oblique axial slices were acquired with an in-plane resolution of 0.9375 mm x 0.9375 mm. Slice thickness and orientation were chosen to be similar to fMRI data for analysis purposes.

The fMRI data were acquired using a one-shot spiral pulse sequence (TR/TE = 2000/35 ms, FOV = 24×24 cm, slice thickness = 3.8 mm, matrix = 64×64). Twenty-six oblique axial slices were acquired with an in-plane resolution of 3.75 mm x 3.75 mm.

Subjects were instructed to perform a single key-press with both index fingers every time they saw the word "tap" appear on a screen. The tap stimulus appeared every 12 seconds and remained on the screen for 1 second. In the interim, subjects were instructed to fixate on a white cross-hair in the middle of the screen. There were 24 trials in a 5-minute block. The cross-hair stimulus appeared every 12 seconds and remained on the screen for 1 second.

FMRI resting-state data were acquired on all 24 subjects; 8 non-depressed subjects (out of 12 total non-depressed comparison subjects) and 11 depressed subjects (out of 12 total depressed subjects) had FLAIR images available for WMH assessment.

This study was approved by the University of Pittsburgh Institutional Review Board and we obtained written informed consent from all study participants.

### 6.2.2 Methods

### 6.2.2.1 WMH burden

An automated WMH segmentation and localization method described in Chapter 5.0 was used to compute the normalized WMH volumes [142]. For each subject, the calculated WMH volume was normalized for the overall brain volume. A natural logarithm transformation was performed on the normalized WMH to minimize the skewness in data.

**6.2.2.2 Resting-state fMRI analyses**

**Preprocessing**. The functional images from each subject were normalized into the standard Montreal Neurological Institute (MNI) template space (Colin27) [143] via the following steps: (i) the functional data were realigned to the first volume using a least square approach and a 6-parameter rigid body transformation to correct for motion correction, (ii) the first volume in the functional data was registered to the subject's high-resolution SPGR image using an affine transformation model, (iii) the subject's SPGR image was then warped to the standard template MNI Colin27 using a deformable model [142], and (iv) the realigned functional images were upsampled and transformed into subject's SPGR space with the affine matrix from step (ii), which were further transformed into the standard space with the computed deformation field from step(iii), (v) the warped functional data were then downsampled to the original resolution of 3.75 mm x 3.75 mm x 3.75 mm; a Gaussian smoothing filter (6 mm full width at half maximum) was then used on the normalized functional images to reduce spatial noise. The fully deformable registration has previously been shown to increase the effect-size with functional imaging [142], and has shown significant improvement in image registration accuracy of aging brains [144].

For each subject, the smoothed, normalized functional images were concatenated to form a 3D+time AFNI dataset. A band-pass temporal filter with the cutoff frequencies of [.01 .1] Hz was then used to extract the resting-state BOLD signal, which also effectively removed the linear trend and high-frequency noise in the data as described in Lowe et al. [145].

**Connectivity Analysis.** Functional connectivity is defined as the temporal correlation of activity between spatially disconnected areas [129], and it has become a popular tool in studying the regional relationships among both healthy individuals and those with clinical disorders [129].

Using this tool, we compared the DMN activity among elderly non-depressed comparison subjects and LLD subjects before and after successful antidepressant treatment.

**Region of Interest (ROI).** The posterior cingulate cortex (PCC) has been shown to have consistently greater activity during resting state than during cognitive tasks [146], and it is hypothesized to constitute a core node in the DMN [127]. Thus, the PCC has often been used as a seed region to identify the DMN. The left and right PCC from the Automated Anatomical Labeling (AAL) atlas [16] (1x1x1 mm) in Colin27 space were down sampled to a voxel resolution of 3.75 mm x 3.75 mm x 3.75 mm (left and right PCC combined, 200 voxels). A smaller region-of-interest (ROI) of 39 voxels, centered on the PCC, was created on the template Colin27 by performing erosions (2 iterations, 6 connected, 2.5-dimensional) with a 3x3x3 (voxel) structuring element. Figure 6.1 presents the original PCC and the eroded ROI centered at PCC on Colin27.
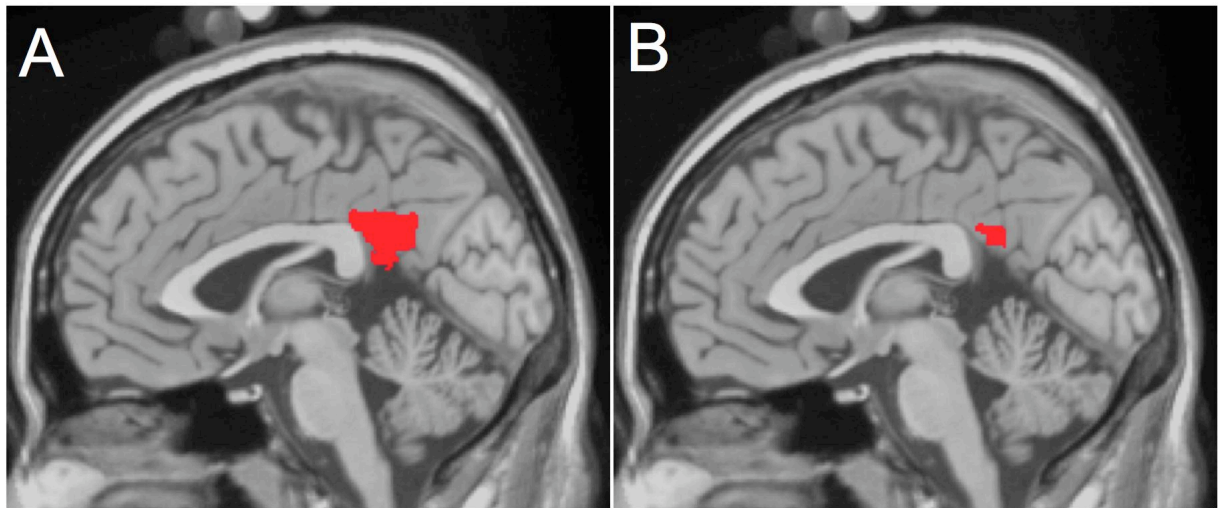


**Figure 6.1. A) The original posterior cingulate cortex (PCC) and B) the eroded PCC are shown in red.**

For each subject, a reference resting-state time-series was extracted by averaging the time-series for all voxels within the PCC ROI. A correlation coefficient map was calculated on

the 3D+time resting state data with the reference time-series as the regressor of interest using 3dDeconvolve from Analysis of Functional NeuroImages (AFNI) [147]. In the map, the correlation coefficient at a given voxel shows the time-series correlation between that voxel and PCC ROI, which represents the resting state functional connectivity score at the voxel.

To determine the mean resting-state functional connectivity map for each group (depressed and comparison subjects), the correlation coefficient maps were statistically compared to the baseline using a 1-sample t-test. The resulting t-maps were then thresholded at a corrected $p < .001$ via Monte Carlo simulations (AlphaSim, AFNI [147], with the whole brain template as mask).

A 2-sample t-test of the correlation coefficient maps (between comparison, pretreatment and post-treatment depressed subjects) was used to identify between-group differences in resting-state functional connectivity. The resulting t-maps were then thresholded at a corrected $p < 0.05$ via Monte Carlo simulations (AlphaSim, AFNI [147], with the ROI mask defined as the conjunctional frontal areas from the compared groups).

### 6.2.2.3 Correlation of WMH Burden and Resting-state Connectivity

The voxel-wise correlation between WMHs and the resting-state functional connectivity score was evaluated using 3dDeconvolve (AFNI) [147]. The functional connectivity correlation coefficient maps from each group (comparison subjects and depressed subjects) were used as the input data and the corresponding log-transformed WMH burden from the same group was used as the regressor of interest in 3dDeconvolve. The F-maps were thresholded at a corrected $p < 0.05$ using Monte Carlo simulations (AlphaSim, AFNI) [147] with the resting-state connectivity map from the related group as ROI mask).

## 6.3 RESULTS

The clinical characteristics of the subjects (depressed and non-depressed comparison subjects) are summarized in Table 1. The difference in Dementia Rating Scale scores between depressed and comparison groups is marginally significant at $p < 0.05$, but becomes non-significant at $p < 0.29$ after adjusting for education level. The CIRC-G cardiovascular and endocrine items did not differ between the two groups.

Table 6.1 Clinical characteristics of the subjects.

| | Group 1 (Non-Depressed) (N=12) | Group 2 (Depressed before treatment) (N=12) | Group 3 (Depression Remitted) (N=8) | Group Difference t-test (t, p) | |
|---|---|---|---|---|---|
| | | | | 1 vs 2 df = 22 | 2 vs 3 df = 6 |
| Subjects (N) | 12 | 12 | ▲8 | | |
| Gender, Female | 4/12 | 7/12 | 3/8 | | |
| Race, Caucasian | •10/12 | 12/12 | 8/8 | | |
| Handedness, Right | 12/12 | 10/12 | 6/8 | | |
| Age, years (mean, SD) | 69.0±6.5 | 70.5±4.9 | 70.8±5.7 | (0.64, 0.53) | |
| Age of onset, years | N/A | 67.9±5.7 | 68.7±6.9 | | |
| °Number (%) of subjects with recurrent depression | N/A | 7/12 (58%) | 4/8 (50%) | | |
| Education, years (mean, SD) | 16.3±2.7 | 13.6±3.0 | 13.4±3.0 | (-2.26, 0.034) | |
| MMSE (mean, SD) | 28.9±1.0 | 28.0±3.5 | 27.8±2.6 | (-0.88, 0.39) | ◆ (0.26, 0.85) |
| DRS (mean, SD) | 140.7±2.8 | 137.0±5.4 | 132.1±9.8 | * (-1.09, 0.29) | △ (2.12, 0.087) |
| HDRS-17 (mean, SD) | 1.5±1.6 | 19.8±4.1 | 6.8±4.5 | (13.69, <.0001) | (5.91, <. 001) |
| CIRS-G (mean, SD) | 5.5±4.6 | 6.3±2.6 | | (0.48, 0.64) | |

Statistical analyses were conducted in SPSS statistical package version 11 for Macintosh.

MMSE: Mini-Mental State Examination; DRS: Dementia Rating Scale; HDRS-17: Hamilton Depression Rating Scale (17 item); CIRS-G: Cumulative Illness Rating Scale for Geriatrics.

## 6.3.1 WMH Burden

The automated WMH segmentation method identified WMHs from the T2-weighted fluid-attenuated inversion recovery (FLAIR) images. The WMH segmentation result for a comparison subject is shown in Figure 6.2. Normalized WMHs (nWMH) for both depressed and comparison groups (comparison group: nWMH mean = 0.53%, SD = 0.28%; depressed group: nWMH mean = 1.50%, SD = 1.76%) showed that the nWMH of the depressed group is significantly higher than the comparison group nWMH (t =1.80, $p < 0.05$) with a one-tailed unequal variance t-test.
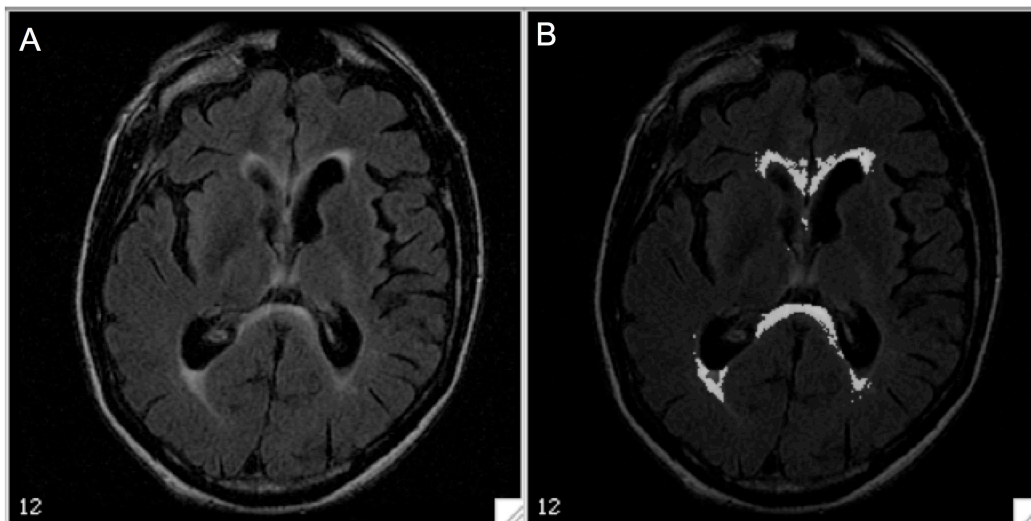


**Figure 6.2. Automated WMH segmentation on the FLAIR image of one subject is shown here, (A) A single slice from a subject's FLAIR image (B) WMH segmentation result with underlying flair image for anatomical reference. The automated WMH segmentation correlates with semi-quantitative, visual rating methods, such**

**as those used for the Cardiovascular Health Study. In the current study, 37% of the depressed subjects had a moderate-severe burden of WMH, as quantified by the Cardiovascular Health Study.**

### 6.3.2   Resting-State Functional Connectivity Maps

Figure 6.3 presents the DMN functional connectivity map from the elderly non-depressed group (Figure 6.3A), and the depressed subject group before treatment (Figure 6.3B) and after 12 weeks of pharmacotherapy (Figure 6.3C). The correlation coefficient maps from each group were statistically compared to the baseline 0 using a 1-sample t-test, and the resulting t-map for the DMN was then thresholded at a corrected $p < .001$ (a joint threshold of $p < 0.01$ and 26-voxel cluster size) via Monte Carlo simulations (AlphaSim, AFNI [147], with whole brain Colin27 as the mask). The standard template MNI colin27 is also displayed as the underlying image for anatomical reference in Figure 6.3.
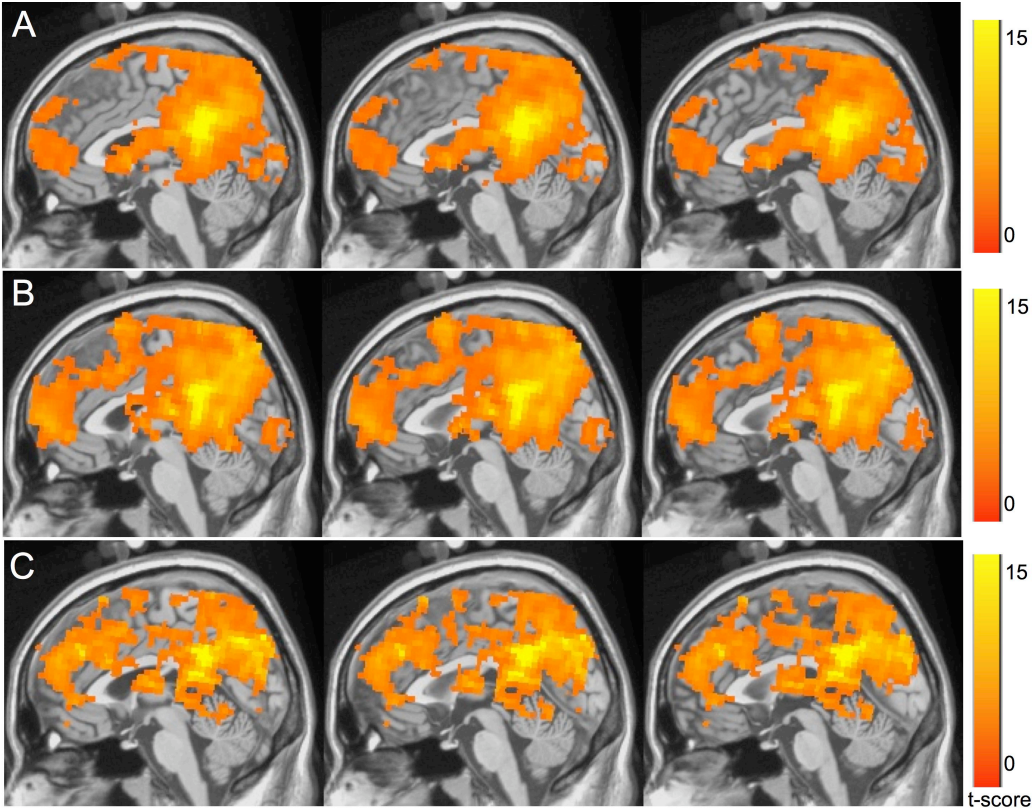
**Figure 6.3. T-maps of the resting-state connectivity for (A) elderly comparison group, late-life depression (LLD) group before treatment (B) and after treatment (C). The maps were thresholded at a corrected p < 0.001. Compared with the elderly comparison group (A), the pretreatment LLD group (B) had significantly lower DMN activation in the subgenual anterior cingulate cortex and significantly higher DMN activation in the dorso-medial prefrontal cortex and the orbito-frontal cortex.**

Figure 6.4A presents the differences in the functional connectivity activities between the comparison group and the pre-treatment depressed group. Figure 6.4B presents the differences in the functional connectivity pattern between patients after treatment and before treatment. Figure 6.4C presents the differences in the functional connectivity pattern between the comparison group and the post-treatment depressed group. In Monte Carlo simulations for between-group comparisons, the ROI mask used is defined as the conjunctional frontal area of DMNs from both groups under comparison.
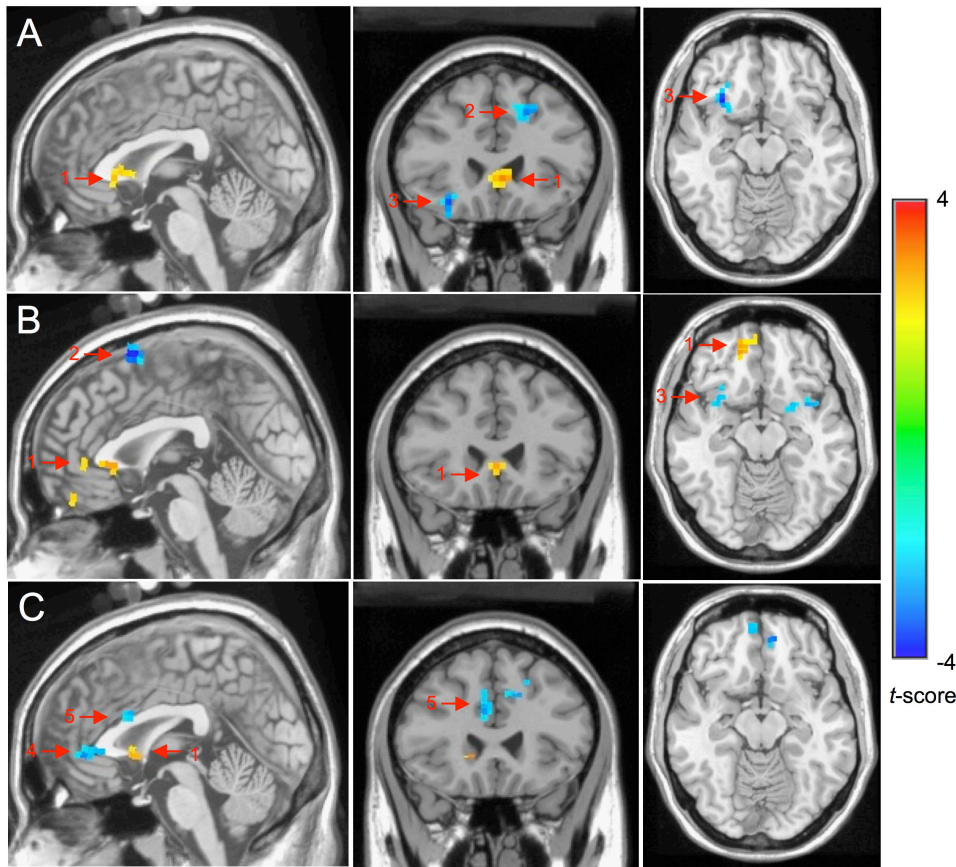
**Figure 6.4. The group comparisons of the functional connectivity activities with two-sample t-test (a corrected p < 0.05) between (A) the comparison group and pretreatment LLD group; (B) pretreatment LLD group and post-treatment LLD group; (C) the comparison group and the post-treatment LLD group. When compared with the elderly comparison group, the pretreatment LLD group had decreased connectivity in the (1) sACC and increased connectivity in the (2) dmPFC and (3) OFC. Compared to pre-treatment LLD group, the post-treatment LLD group has improved connectivity in sACC and dmPFC. Compared with the non-depressed group, the post-treatment LLD group has decreased connectivity in (1) sACC and increased connectivity in (4) rostral ACC and (5) dorsal ACC.**

From the above analyses, we found that:

(1) Compared to non-depressed elderly subjects, elderly depressed subjects pretreatment had significantly <u>lower</u> DMN activation in the BA25 subgenual anterior cingulate cortex (sACC; Talairach coordinates x=81, y=60, z=64; corrected p < 0.05) and significantly <u>higher</u> DMN

activation in the BA 6 dorso-medial prefrontal cortex (dmPFC; Talairach coordinates x=93, y=62, z=106; corrected p < 0.05) and the orbito-frontal cortex (OFC; Talairach coordinates x=52, y=60, z=57; corrected p < 0.05). For all these regions we used a joint threshold of p < 0.05 and 21-voxel cluster size (see Figure 6.4A). In Monte Carlo simulations, the ROI mask is defined as the conjunctional frontal area of DMNs from non-depresssed and depressed elderly subjects.

(2) Compared to themselves before treatment, depressed subjects who responded to 12 weeks of pharmacotherapy exhibited significant improvement (i.e. normalization) in the resting-state connectivity in both sACC (corrected p < 0.05; joint threshold of p < 0.05 and 24-voxel cluster size, using AlphaSim with the defined ROI mask) and dmPFC (corrected p < 0.05) (see Figure 6.4B).

(3) Compared with non-depressed elderly subjects, post-treatment elderly depressed subjects had significantly lower DMN activation in the sACC (corrected p < 0.05; joint threshold of p< 0.05 and 21-voxel cluster size, using AlphaSim with the defined ROI mask) and a higher DMN activation in the rostral ACC (Talairach coordinates x=79, y=46, z=72; corrected p < 0.05) and dorsal ACC (Talairach coordinates x=79, y=69, z=89; corrected p < 0.05) (see Figure 6.4C).

### 6.3.3   Correlation of WMH Burden and Resting-state Connectivity

In LLD subjects pretreatment, there was a significant negative correlation between log-transformed nWMH volume and resting state connectivity (N= 11, averaged $r$ = -0.72, corrected $p$ < 0.05; joint threshold of p < 0.05 and 58-voxel cluster size) in the medial frontal region, i.e., higher WMH volume was associated with lower resting state connectivity.  The r and t maps are shown in Figure 6.5. In Monte Carlo simulations, the ROI mask used is defined as the resting-state DMN from the LLD group (see Figure 6.3B).

There was a significant negative correlation between log-transformed nWMH volume and resting state connectivity in elderly non-depressed comparison subjects in the medial prefrontal region (N = 8; averaged r = -0.80; corrected p < 0.05; joint threshold of p < 0.05 and 58-voxel cluster size). In Monte Carlo simulations, the ROI mask used is defined as the resting-state DMN from the non-depressed group (see Figure 6.5A). With the assumption that the WMH remained unchanged after 12 weeks of treatment, WMH was measured only pretreatment in the depressed group. Thus, we cannot make inferences regarding the correlation of WMH and resting state scores of post-treatment depressed subjects.
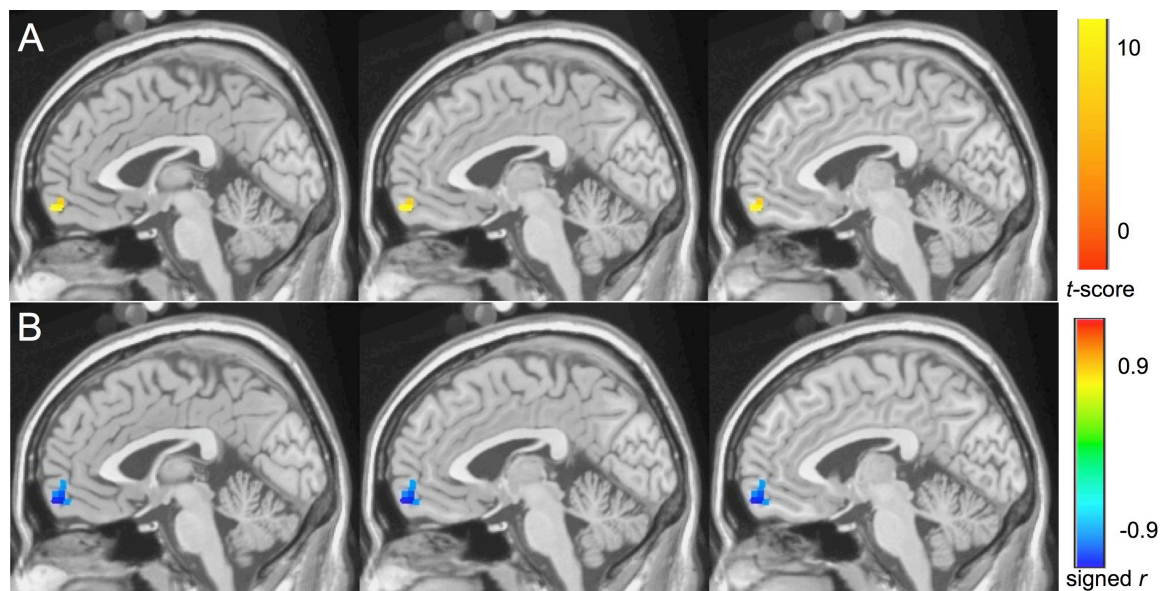


**Figure 6.5. The WMH burden and resting state score for the pretreatment LLD group shows a significant negative correlation (averaged r =-0.72, corrected p < 0.05) in the medial frontal region: (A) t map, (B) signed r map.**

## 6.4    DISCUSSION AND SUMMARY

To our knowledge, this is the first study describing the functional connectivity of the DMN in LLD. Our results show abnormal connectivity patterns in the prefrontal branch of the DMN in acutely depressed elderly subjects. This abnormal functional connectivity is significantly correlated with greater WMH volume. Moreover, we found some improvement in functional connectivity of DMN following treatment responses. We also confirm previous results regarding increased burden of WMH in LLD subjects when compared to non-depressed individuals.

### 6.4.1    Altered Resting State Connectivity in LLD

Previous reports regarding the DMN functional connectivity in midlife depression have reported increased subgenual ACC connectivity (sACC) [130]. Our study shows that subjects with LLD have overlapping but somewhat different functional connectivity of the DMN than midlife MDD subjects. The sACC-PCC connectivity was decreased in acutely depressed elderly, and although it improved after treatment response, post-treatment LLD subjects continued to present decreased sACC-PCC connectivity when compared to non-depressed control subjects. However, different data collection and data analysis methods were used in the midlife depression resting-state studies [130]. These differences may impede the comparison of the patterns of activation in midlife and late-life depression.

Our results emphasize the importance of sACC as an impaired hub in the functional neuroanatomy of MDD [148, 149]. Our results also suggest the importance of the cerebrovascular component of LLD represented by the increased WMH burden.

Two areas had increased functional connectivity in the elderly subjects: the dmPFC and the OFC (see Figure 6.4). With regards to the dmPFC-PCC increased functional connectivity in LLD, while DMN areas are generally attenuated by effortful cognitive tasks, some studies have described a split-effect of different stimuli on DMN activation. Thus, increased self-referential activities (such as judgments about one's mental or emotional state) are less deactivating or even specifically activate the dmPFC in contrast with other DMN areas; these changes have been interpreted as a biological marker of increased negative self-attributional style of depressed subjects [130, 150, 151]. Based on these reports, one may speculate that the observed OFC hyperactivity may be related to difficulties reappraising information – translated into hopelessness and decision-making difficulties - which may be associated with the cognitive changes described in LLD [152].

### 6.4.2 Evidence of "vascular depression" hypothesis of LLD

The correlation between WMH volume and dysfunctional connectivity in the DMN supports the "vascular depression" hypothesis of LLD [123], even though the multifactorial etiology of WMH (including demelination due to neurodegeneration) does not allow for a definitive conclusion. The post-treatment persistence of sACC-PCC hypo-connectivity suggests a possible persistent feature of LLD, perhaps associated with the structural brain changes.

### 6.4.3 Advantages and Limitations

Beyond the novel subject population, our study has several other strengths: (1) a comparison group, which allowed us to compare LLD pre- and post-treatment to non-depressed elders; (2) an automated WMH segmentation method, which provided an objective, quantitative assessment of

WMH volume; (3) a fully deformable model for inter-subject image normalization, which has previously been shown to significantly improve co-localization of the fMRI signal in aging brains; and (4) acquisition of data pre- and post-treatment, allowing us to make inferences regarding the persistent biological features of LLD.

Our report also has some limitations: we had a relatively small sample, especially in the post-treatment group and this limitation hampered our ability to analyze possible differences between LLD responders and non-responders. Our ability to detect WMH burden differences was hampered by the lack of FLAIR images for some of our subjects. Since, we computed the whole brain WMH burden, thus we could not make more specific inferences regarding the role of region-specific WMH burden and the decreased functional connectivity in the DMN. Also, with regard to WMH burden, the significant group difference between LLD and controls was lost when a one-tailed t-test was changed to a two-tailed test. Although none of the participants had evidence of stroke, we did not quantify the presence of lacunae. Overall, depressed subjects had less education than the comparison group, however this had no effect on the pre-post treatment findings in depressed subjects. As there was no significant difference between depressed subjects and non-depressed comparison subjects with regards to both MMSE and DRS, it is unclear if and how the lower education levels of the depressed subjects interfered with the fMRI findings. Given the design of the parent study, we have no imaging data on the depressed non-responders, as they were excluded from the study. Thus, we cannot make further inferences regarding the correlation between DMN changes and treatment response. The between-group comparisons were threshold at a corrected $p<0.05$; the lower significance threshold allows for a greater risk for type I error. However, we used the Monte Carlo simulations (AFNI) to correct for multiple comparisons.

In conclusion, we present novel data highlighting some of the biological basis of LLD. Given the high rate of WMH and concordant cerebrovascular disease, further research may tackle the treatment advantages of using adjuvant calcium channel blockers in the treatment of LLD [153]. Future research is also necessary to understand the correlation between DMN changes and various clinical characteristics of LLD, such as severity, duration of current episode, recurrence, comorbidity with anxiety or medical conditions, and treatment response.

# 7.0    CONCLUSIONS

Heterogeneous changes in the human brain during aging have presented a significant challenge in analyzing MR brain images of the elderly. Methods developed in this dissertation have been shown to improve the accuracy of registration and segmentation for the elderly MR brain images. The key contribution is in describing and testing specific models that can characterize the aged-related changes in the elderly brains. For example, a deformable registration pipeline allows for the highly nonlinear differences present in elderly brains, and thus provides more accurate deformation estimation for image registration of elderly brain images.  Similarly, the optimum template strategy and the multi-template strategy both can better represent the heterogeneous changes in the elderly brains, which lead to improved registration accuracy in atlas-based segmentation and image normalization, respectively.

Experiments have been designed to evaluate the performance characteristics of the developed methods, compared to the traditional method. Results from these experiments were encouraging and have shown improved performance with the developed methods. Some of the results from the designed experiments were intriguing as well. For example, in Chapter 2.0 lower segmentation accuracy was found for the anterior cingulate cortex (ACC) than for the hippocampus in atlas-based segmentation using the deformable registration. This observation then intrigued us to suspect that the lower accuracy for the ACC was probably due to the inter-subject variability in the gyral folding pattern in this structure. Therefore, the optimum template

selection strategy was proposed and has shown significantly improved segmentation accuracy for ACC using a set of 3 prototype templates with wide anatomical variations. Some of the results were puzzling, for which possible reasons and potential future directions were suggested (see details in Chapter 7.2).

The remainder of this chapter summaries the proposed methods and the experimental results, and then discusses potential future directions for neuroimage research in the elderly population.

## 7.1    GOALS REVISITED

The motivations and goals of this dissertation presented in Chapter 1.3 are revisited and reviewed in this section.

**Goal 1** Implementation of a fully deformable registration pipeline and quantitative comparison of the deformable model and AIR and SPM.

A fully deformable registration has been developed using the registration library from ITK. A series of experiments have been designed for quantitative comparison of the deformable model, AIR and SPM. Results showed that the performance of the deformable model was significantly better than AIR and SPM in the atlas-based segmentation, image normalization as well as the co-localization of fMRI signal.

**Goal 2** An optimum template strategy for atlas-based segmentation.

Given the normal and age-related variations in the brain structures of the elderly population, an optimum template strategy has been proposed and developed to improve the

segmentation accuracy in the atlas-based segmentation of elderly MR brain images. Results showed that, compared to the traditional single template method, the developed optimum template method provided significantly superior segmentation results for the most brain regions. Results also suggested that in addition to template number, variability in the selected family of templates also played an important role in improving the performance of the optimum template. We have shown that a small number of templates with appropriate anatomical variations were sufficient to achieve high accuracy in the atlas-based segmentation.

**Goal 3** A multi-template strategy for image normalization

A novel multi-template strategy has been proposed and validated in this dissertation. In this method, intermediate prototype templates were used as a bridge in image normalization, and the registration was carried out as subject->intermediate brain images->Colin27. Results showed that the multi-template strategy (subject→s-image→r-image→Colin27) significantly improved the registration accuracy for linear and nonlinear registration algorithms. This multi-template method can better characterize the anatomical variations in the elderly population, and at the same time retains the detailed structures in the templates. Part of the registration pipeline (s-image→r-image→Colin27) can be computed offline to take advantage of computationally expensive registration methods.

**Goal 4** Segmentation and localization of white matter hyperintensities

The fuzzy connected algorithm with automated seed identification was proposed and developed for WMH segmentation. A registration pipeline using the deformable model was developed for WMH localization. Compared to the traditional visual grading methods, this method provided more accurate WMH volume measurement and more specific information on

112

the spatial distribution of the WMHs, and thus it was able to detect subtle WMH differences between groups.

**Goal 5** Resting state connectivity and white matter burden in late life depression

With the developed methods (the deformable model and the WMH segmentation method), the resting state connectivity and the effect of white matter hyperintensity were studied in late-life depression. In this clinical application, abnormal connectivity patterns were discovered in the prefrontal area of the default-mode network in depressed elderly subjects and this abnormal functional connectivity was significantly correlated with greater WMH volume. These results provided evidence for the "vascular depression" hypothesis of late life depression.

## 7.2    FUTURE RESEARCH AND DIRECTIONS

In this dissertation, the deformable model has been developed to improve the registration accuracy of structural and functional MR brain images in the elderly population. The research on structural images mainly focuses the overlap and volume brain structures. However, volume measurement of the brain structures may not be the best way to characterize the subtle gray matter changes in neurodegenerative and psychiatric disorder. Cortical thickness can estimate the gray matter changes with more regional specificity and better precision (submillimeter), which makes it particular useful in detecting the subtle cortical atrophy at the early stage of neurodegenerative diseases [154]. It will be interesting to develop a cortical thickness method based on the deformable registration model. This method may be able to provide more accurate thickness measurement of the gray matter, given the improved registration accuracy from the deformable model.

Results from Chapter 3.0 showed that the optimum template selection has produced improved anatomical segmentation accuracy for all the analyzed ROIs except for the left and right amygdala. In this case, the deformable model failed to provide an accurate segmentation for amygdala with the template candidates. Possible reasons for the unsuccessful registration are: 1) the deformable registration model is not good enough in amygdala segmentation. Alternative segmentation methods (perhaps using feature-based registration) may improve the automated labeling of the small and difficult to segment regions such as the left and right amygdala. 2) the random selected template candidates do not represent the variability in the amygdala. Further research is needed to explore the normal variations of the brain anatomy.

Chapter 3.0 also showed that the number of templates as well as the anatomical varibilities in the templates could affect the performance of the optimum template method. A subset of the templates with wide anatomical variations was found to be sufficient to provide improved ACC segmentation. However, the anatomical variations in other brain structures have not been studied. For future research direction, machine learning algorithms or shape morphometry analyses could be combined or used independently to identify sufficient number of templates for brain structures of interest.

The multi-template strategy (subject→intermediate brain images→Colin27) has shown superior accuracy in image normalization when using linear and nonlinear registration algorithms. However, contrary to our prediction, the multi-template strategy did not improve the image normalization when using the deformable model. Possible reasons have also been explored in Section 4.5. Improved registration accuracy between the training data and the MNI template Colin27 (intermediate brain images→Colin27) will improve overall image normalization. The intermediate brain images→Colin27 registration can be computed offline and

the obtained deformation fields will be repeatedly used in future studies. To take advantage of this feature, in future work, a more computationally expensive registration algorithms or a landmark based registration method, or an iterative combination of both methods could be used to provide a nearly perfect registration between the training data and the MNI template Colin27.

An automated method has been developed in Chapter 5.0 to detect the abnormal white matter in T2-weighted FLAIR images. This method provides more accurate and more regional specific measurement of white matter hyperintensities. However, not all clinical studies include T2-weighted FLAIR image in their scanning protocols, but most studies have basic T1-weighted and T2-weighted brain images. Future research includes a modification of the current method to automatically segment the WMH on T2-weighted images. Different criteria for automated seed identification will be explored on T2-weighted image. Images with different modalities can help to improve the seed identification, such as T1-weighted image is used to mask out the bright CSF on T2-weighted image and etc.

In the clinical application, the effect of overall white matter burden on resting state connectivity was studied. Further multi-modality MR studies can provide additional evidence of the relationship between structural and functional alterations in brain connectivity in late-life depression. These future studies may provide more specific evidence for the "vascular depression" hypothesis of late life depression.

# BIBLIOGRAPHY

1. Kikinis, R., et al., *Computer-assisted interactive three-dimensional planning for neurosurgical procedures.* Neurosurgery, 1996. **38**(4): p. 640-9; discussion 649-51.

2. Ogawa, S., et al., *Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging.* Proc Natl Acad Sci U S A, 1992. **89**(13): p. 5951-5.

3. Pelizzari, C.A., et al., *Accurate three-dimensional registration of CT, PET, and/or MR images of the brain.* J Comput Assist Tomogr, 1989. **13**(1): p. 20-6.

4. Woods, R.P., et al., *Automated image registration: II. Intersubject validation of linear and nonlinear models.* J Comput Assist Tomogr, 1998. **22**(1): p. 153-65.

5. Collins, D.L., et al., *Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space.* J Comput Assist Tomogr, 1994. **18**(2): p. 192-205.

6. Friston, K.J., et al., *Spatial registration and normalization of images.* Human Brain Mapping, 1995. **3**(3): p. 165-189.

7. Woods, R.P., et al., *Automated image registration: I. General methods and intrasubject, intramodality validation.* J Comput Assist Tomogr, 1998. **22**(1): p. 139-52.

8. Gee, J.C., M. Reivich, and R. Bajcsy, *Elastically deforming 3D atlas to match anatomical brain images.* J Comput Assist Tomogr, 1993. **17**(2): p. 225-36.

9. Collins, D.L., et al., *Automatic 3-D model-based neuroanatomical segmentation.* Human Brain Mapping, 1995. **3**: p. 190-208.

10. Christensen, G.E., R.D. Rabbitt, and M.I. Miller, *3D brain mapping using a deformable neuroanatomy.* Phys Med Biol, 1994. **39**(3): p. 609-18.

11. Thirion, J.P., *Image matching as a diffusion process: an analogy with Maxwell's demons.* Med Image Anal, 1998. **2**(3): p. 243-60.

12. Kim, B., et al., *Mutual information for automated unwarping of rat brain autoradiographs.* Neuroimage, 1997. **5**(1): p. 31-40.

13. Kikinis, R., et al., *A Digital Brain Atlas for Surgical Planning, Model-Driven Segmentation, and Teaching.* IEEE Transactions on Visualization and Computer Graphics, 1996. **2**(3): p. 232-241.

14. Toga, A.W., *Brain warping.* 1999, San Diego: Academic Press. xiii, 385 p.

15. Holmes, C.J., et al., *Enhancement of MR images using registration for signal averaging.* Journal of Computer Assisted Tomography, 1998. **22**(2): p. 324-333.

16. Tzourio-Mazoyer, N., et al., *Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.* Neuroimage, 2002. **15**(1): p. 273-89.

17. Paus, T., et al., *Human cingulate and paracingulate sulci: pattern, variability, asymmetry, and probabilistic map.* Cereb Cortex, 1996. **6**(2): p. 207-14.

18. Resnick, S.M., et al., *Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain.* J Neurosci, 2003. **23**(8): p. 3295-301.

19. Raz, N., et al., *Regional brain changes in aging healthy adults: general trends, individual differences and modifiers.* Cereb Cortex, 2005. **15**(11): p. 1676-89.

20. Walhovd, K.B., et al., *Effects of age on volumes of cortex, white matter and subcortical structures.* Neurobiol Aging, 2005. **26**(9): p. 1261-70; discussion 1275-8.

21. Woods, R.P., et al., *Automated image registration: II. Intersubject validation of linear and nonlinear models.* Journal of Computer Assisted Tomography, 1998. **22**(1): p. 153-65.

22. Talairach, J. and P. Tournoux, *Co-planar stereotaxic atlas of the human brain : 3-dimensional proportional system: an approach to medical cerebral imaging.* 1988, Stuttgart: Thieme Medical Publishers. 122.

23. Sneath, P.H.A., *The application of computers to taxonomy.* J. Gen. Microbiol., 1957. **17**: p. 201–226.

24. Jian, A. and R. Dubes, *Algorithms for Clustering Data.* 1988, Englewood

Cliffs, NJ: Prentice-Hall.

25. Johnson, S.C., *Hierarchical clustering schemes.* Psychometrika, 1967. **32**: p. 241-254.

26. Everitt , B., S. Landau, and M. Leese, *Cluster Analysis.* 2001, London: Arnold.

27. van Gijn, J., *Leukoaraiosis and vascular dementia.* Neurology, 1998. **51**(3 Suppl 3): p. S3-8.

28. Mirsen, T., et al., *Clinical correlates of white-matter changes on magnetic resonance imaging scans of the brain.* Arch Neurol, 1991. **48**: p. 1015-1021.

29. Hickie, I. and E. Scott, *Late-onset depressive disorders: a preventable variant of cerebrovascular disease?* Psychological Medicine, 1998. **28**(5): p. 1007-1013.

30. Alan J. Thomas, Rajesh N. Kalaria, and J.T. O'Brien, *Depression and vascular disease: what is the relationship?* Journal of Affective Disorders, 2004. **79**(1-3): p. 81-95.

31. Yoo, T.S., *Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis.* 2004, Wellesey, MA: AK Peters Ltd.

32. Chen, M., *3-D deformable registration using a statistical atlas with applications in medicine*, in *Robotics Institute Technical Report. Pittsburgh, PA.* 1999, Carnegie Mellon University.

33. Carmichael, O.T., et al., *Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment.* Neuroimage, 2005. **27**(4): p. 979-90.

34.   Lopez-Garcia, P., et al. *Automatic MRI-based brain parcellation and regional volume measurement: a validation study in healthy subjects and individuals with schizophrenia.* in *Human Brain Mapping.* 2003. N.Y.

35.   Smith, S.M., *Fast robust automated brain extraction.* Hum Brain Mapp, 2002. **17**(3): p. 143-55.

36.   Rosano, C., et al., *Morphometric analysis of gray matter volume in demented older adults: exploratory analysis of the cardiovascular health study brain MRI database.* Neuroepidemiology, 2005. **24**(4): p. 221-9.

37.   Cox, R.W., *AFNI: software for analysis and visualization of functional magnetic resonance neuroimages.* Comput Biomed Res, 1996. **29**(3): p. 162-73.

38.   Xiong, J.H., et al., *Clustered pixels analysis for functional MRI activation studies of the human brain.* Human Brain Mapping, 1995. **3**(4): p. 287-301.

39.   Carter, C.S., et al., *Anterior cingulate cortex, error detection, and the online monitoring of performance.* Science, 1998. **280**(5364): p. 747-9.

40.   Clark, K.A., et al. *Conflict Monitoring in the absence of awareness: an event-related fMRI study.* in *Society for Neuroscience.* 2002. Orlando, FL.

41.   Carter, C.S., et al., *Parsing executive processes: strategic vs. evaluative functions of the anterior cingulate cortex.* Proc Natl Acad Sci U S A, 2000. **97**(4): p. 1944-8.

42.   Wu, M., C. Rosano, and H.J. Aizenstein, *A morphological method to improve skull stripping of MR brain images.* , in *the 11th Annual Meeting of the Organization for Human Brain Mapping.* 2005: Toronto, Canada.

43.   Wu, M., et al., *Quantitative comparison of AIR, SPM, and the fully deformable model for atlas-based segmentation of functional and structural MR images.* Hum Brain Mapp, 2006. **27**(9): p. 747-54.

44.   Carmichael, O.T., et al., *Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment* NeuroImage, 2005. **27**(4): p. 979-990.

45.   Vemuri , B.C., et al., *Image registration via level-set motion: Applications to atlas-based segmentation.* Medical Image Analysis, 2003. **7**(1): p. 1-20.

46.   Dawant, B.M., et al., *Automatic 3-D Segmentation of Internal Structures of the Head in MR Images Using a Combination of Similarity and Free-Form Transformations: Part I, Methodology and Validation on Normal Subjects.* IEEE Transactions on Medical Imaging, 1999. **18**(10): p. 909 - 916.

47.   Thompson, P.M., et al., *Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain.* Hum Brain Mapp, 2000. **9**(2): p. 81-92.

48.   Sabuncu, M.R., S.K. Balci, and P. Golland, *Discovering modes of an image population through mixture modeling.* Medical Image Computing and Computer-Assisted Intervention - Miccai 2008, Pt 2, 2008. **5242**: p. 381-389.

49.   Collins, D.L., *3D Model-based segmentation of individual brain structures from magnetic resonance imaging data*, in *Department of Biomedical Engineering.* 1994, McGill University: Montreal.

50. Rex, D.E., J.Q. Ma, and A.W. Toga, *The LONI Pipeline Processing Environment.* Neuroimage, 2003. **19**(3): p. 1033-48.

51. Rohlfing, T., et al., *Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains.* Neuroimage, 2004. **21**(4): p. 1428-42.

52. Rohlfing, T., D.B. Russakoff, and C.R. Maurer, Jr., *Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation.* IEEE Trans Med Imaging, 2004. **23**(8): p. 983-94.

53. Heckemann, R.A., et al., *Automatic anatomical brain MRI segmentation combining label propagation and decision fusion.* Neuroimage, 2006. **33**(1): p. 115-26.

54. Wu, M., et al., *Optimum template selection for atlas-based segmentation.* Neuroimage, 2007. **34**(4): p. 1612-8.

55. Aljabar, P., et al., *Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy.* Neuroimage, 2009. **46**(3): p. 726-38.

56. Rohlfing, T., et al., *Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains.* NeuroImage, 2004. **21**(4): p. 1428-1442.

57. Chen, M., *3-D deformable registration using a statistical atlas with applications in medicine.* 1999, Carnegie Mellon University.

58. Unser, M., A. Aldroubi, and C.R. Gerfen. *A Multiresolution Image Registration Procedure Using Spline Pyramids.* in *Proceedings of the SPIE Conference on Mathematical Imaging: Wavelet Applications in Signal and Image Processing.* 1993. San Diego CA, USA.

59. Thirion, J.P., *Image matching as a diffusion process: an analogy with Maxwell's demons.* Medical Image Analysis, 1998. **2**: p. 243-60.

60. Yoo, T., *Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis.* 2004, Wellesey, MA: AK Peters Ltd.

61. Dice, L.R., *Measures of the amount of ecologic association between species.* Ecology, 1945. **26**: p. 297–302

62. Venkatasubramanian, G., et al., *Striato-cerebellar abnormalities in never-treated schizophrenia.* Biological Psychiatry, 2003. **53**(8): p. 23S-23S.

63. Michio Ono, S. Kubik, and Chad D. Abernathey, *Atlas of the cerebral sulci.* 1990, Stuttgart: New York: G. Thieme Verlag.

64. Ashburner, J. and K.J. Friston, *Spatial Normalization*, in *Brain Warping*, A.W. Toga, Editor. 1998, Academic Press.

65. Toews, M., D.L. Collins, and T. Arbel, *A statistical parts-based appearance model of inter-subject variability.* Medical Image Computing and Computer-Assisted Intervention - Miccai 2006, Pt 1, 2006. **4190**: p. 232-240.

66.	Gur, R.C., et al., *Gender differences in age effect on brain atrophy measured by magnetic resonance imaging.* Proc Natl Acad Sci U S A, 1991. **88**(7): p. 2845-9.

67.	Greenberg, D.L., et al., *Aging, gender, and the elderly adult brain: an examination of analytical strategies.* Neurobiol Aging, 2008. **29**(2): p. 290-302.

68.	Park, H., et al., *Least biased target selection in probabilistic atlas construction.* Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv, 2005. **8**(Pt 2): p. 419-26.

69.	Marsland, S., C.J. Twining, and C.J. Taylor, *Groupwise non-rigid registration using polyharmonic clamped-plate splines.* Medical Image Computing and Computer-Assisted Intervention - Miccai 2003, Pt 2, 2003. **2879**: p. 771-779.

70.	Geng, X., et al., *Implicit reference-based group-wise image registration and its application to structural and functional MRI.* Neuroimage, 2009. **47**(4): p. 1341-51.

71.	Studholme, C. and V. Cardenas, *A template free approach to volumetric spatial normalization of brain anatomy.* Pattern Recognition Letters, 2004. **25**(10): p. 1191-1202.

72.	Joshi, S., et al., *Unbiased diffeomorphic atlas construction for computational anatomy.* Neuroimage, 2004. **23 Suppl 1**: p. S151-60.

73.	Tang, S., et al., *RABBIT: rapid alignment of brains by building intermediate templates.* Neuroimage, 2009. **47**(4): p. 1277-87.

74.	Gauthier, S., et al., *Mild cognitive impairment.* Lancet, 2006. **367**(9518): p. 1262-70.

75.	Petersen, R.C., et al., *Mild cognitive impairment: clinical characterization and outcome.* Arch Neurol, 1999. **56**(3): p. 303-8.

76.	Jack, C.R., Jr., et al., *The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods.* J Magn Reson Imaging, 2008. **27**(4): p. 685-91.

77.	Hellier, P., et al., *Inter-subject registration of functional and anatomical data using SPM.* Medical Image Computing and Computer-Assisted Intervention - Miccai 2002, Pt 2, 2002. **2489**: p. 590-597.

78.	Xiong, J., et al., *Clustered pixels analysis for functional MRI activation studies of the human brain.* Human Brain Mapping, 1995. **3**: p. 287-301.

79.	Dale, A.M., B. Fischl, and M.I. Sereno, *Cortical surface-based analysis. I. Segmentation and surface reconstruction.* Neuroimage, 1999. **9**(2): p. 179-94.

80.	Fischl, B., M.I. Sereno, and A.M. Dale, *Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system.* Neuroimage, 1999. **9**(2): p. 195-207.

81.	Thomas, A.J., R.N. Kalaria, and J.T. O'Brien, *Depression and vascular disease: what is the relationship?* Journal of Affective Disorders, 2004. **79**(1-3): p. 81-95.

82.	Udupa, J.K., P.K. Saha, and R.A. Lotufo, *Relative fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation.* Ieee Transactions on Pattern Analysis and Machine Intelligence, 2002. **24**(11): p. 1485-1500.

83.     Udupa, J.K. and S. Samarasekera, *Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation.* Graphical Models and Image Processing, 1996. **58**(3): p. 246-261.

84.     O'Brien, J., D. Ames, and I. Schwietzer, *White matter changes in depression and Alzheimer's disease: a review of magnetic resonance imaging studies.* Int J Geriatr Psychiatry, 1996. **11**: p. 681-694.

85.     Bryan, R., et al., *A method for using MR to evaluate the effects of cardiovascular disease on the brain: the Cardiovascular Health Study.* Am J Neuroradiol, 1994. **15**: p. 1625-1633.

86.     Yue, N., et al., *Sulcal, ventricular, and white matter changes at MR imaging in the aging brain: data from the Cardiovascular Health Study.* Radiology, 1997. **202**: p. 33-39.

87.     Fazekas, F., et al., *MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging.* American Journal of Roentgenology, 1987. **149**(2): p. 351-6.

88.     Davis, P.C., et al., *The consortium to establish a registry for Alzheimer's disease (CERAD). Part III. reliability ofa standardized MRI evaluation of Alzheimer's disease.* Neurology, 1992. **42**(9): p. 1676-1680.

89.     Anbeek, P., et al., *Automatic segmentation of different-sized white matter lesions by voxel probability estimation.* Medical Image Analysis, 2004. **8**(3): p. 205-215.

90.     Anbeek, P., et al., *Probabilistic segmentation of white matter lesions in MR imaging.* Neuroimage, 2004. **21**(3): p. 1037-1044.

91.     Swartz, R.H., et al., *Utility of simultaneous brain, CSF and hyperintensity quantification in dementia.* Psychiatry Research Neuroimaging, 2002. **116**: p. 83-93.

92.     Kikinis, R., et al., *Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging.* Journal of Magnetic Resonance Imaging., 1992. **2**: p. 619-629.

93.     Pachai, C., et al., *A pyramidal approach for automatic segmentation of multiple sclerosis lesions in brain MRI.* Computerized Medical Imaging and Graphics., 1998. **22**: p. 399-408.

94.     Stamatakis, E.A. and L.K. Tyler, *Identifying lesions on structural brain images-Validation of the method and application to neuropsychological patients.* Brain and Language, 2005. **94**: p. 167-177.

95.     Hirono, N., et al., *Impact of white matter changes on clinical manifestation of Alzheimer's disease a quantitative study. Stroke*, 2000. **31**: p. 2182 - 2188.

96.     Jack, C.R., Jr., , et al., *FLAIR histogram segmentation for measurement of leukoaraiosis volume.* Journal of Magnetic Resonance Imaging, 2001. **14**(6): p. 668-676.

97.     Wen, W. and P. Sachdev, *The topography of white matter hyperintensities on brain MRI in healthy 60- to 64-year-old individuals.* Neuroimage, 2004. **22**(1): p. 144-154.

98.     Benson, R.R., et al., *Older people with impaired mobility have specific loci of periventricular abnormality on MRI.* Neurology, 2002. **58**: p. 48-55.

99.    Taylor, W.D., et al., *Localization of age-associated white matter hyperintensities in late-life depression.* Progress in Neuro-Psychopharmacology & Biological Psychiatry., 2003. **27**(3): p. 539-44.

100.   DeCarli, C., et al., *Anatomical mapping of white matter hyperintensities (WMH): exploring the relationships between periventricular WMH, Deep WMH, and total WMH burden.* Stroke., 2005. **36**: p. 50-55.

101.   Association, A.P., *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR.* American Psychiatric Association, 2000.

102.   Beekman, A.T., et al., *Anxiety and depression in later life: Co-occurrence and communality of risk factors.* American Journal of Psychiatry, 2000. **157**(1): p. 89-95.

103.   Smith, S.M., *Fast robust automated brain extraction.* Human Brain Mapping, 2002. **17**(3): p. 143-55.

104.   Udupa, J.K. and S. Samarasekera *Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation.* Graphical Models and Image Processing, 1996. **58**(3): p. 246-261.

105.   Udupa, J.K., et al., *Multiple sclerosis lesion quantification using fuzzy-connectedness principles.* IEEE Transactions on Medical Imaging, 1997. **16**(5): p. 598-609.

106.   Rosano, C., et al., *Morphometric Analysis of Gray Matter Volume in Demented Older Adults: Exploratory Analysis of the Cardiovascular Health Study Brain MRI Database.* Neuroepidemiology, 2005. **24**(4): p. 221-9.

107.   Aizenstein, H.J., et al., *Prefrontal and Striatal Activation During Sequence Learning in Geriatric Depression.* Biological Psychiatry, In Press.

108.   Wu, M., et al., *Quantitative Comparison of Neuroimage Registration by AIR, SPM, and a Fully Deformable Model.* Human Brain Mapping, In Press.

109.   Wakana, S., et al., *Fiber tract–based atlas of human white matter anatomy.* Radiology, 2004. **230**: p. 77-87.

110.   Schmidt, R., et al., *White matter lesion progression, brain atrophy, and cognitive decline: the Austrian stroke prevention study.* Annals of Neurology, 2005. **58**(4): p. 610-616.

111.   Capizzano, A.A., et al., *White matter hyperintensities are significantly associated with cortical atrophy in Alzheimer's disease.* Journal of Neurology Neurosurgery and Psychiatry, 2004. **75**: p. 822-827.

112.   Miki, Y., et al., *Computer-assisted quantitation of enhancing lesions in multiple sclerosis: correlation with clinical classification.* American Journal of Neuroradiology, 1997. **18**(4): p. 705-710.

113.   Charney, D.S., et al., *Depression and Bipolar Support Alliance consensus statement on the unmet needs in diagnosis and treatment of mood disorders in late life.* Arch Gen Psychiatry, 2003. **60**(7): p. 664-72.

114.   Greenwald, B.S., et al., *Qualitative magnetic resonance imaging findings in geriatric depression. Possible link between later-onset depression and Alzheimer's disease?* Psychol Med, 1997. **27**(2): p. 421-31.

115.    Ballmaier, M., et al., *Localizing gray matter deficits in late-onset depression using computational cortical pattern matching methods.* Am J Psychiatry, 2004. **161**(11): p. 2091-9.

116.    Steffens, D.C., et al., *Hippocampal volume in geriatric depression.* Biol Psychiatry, 2000. **48**(4): p. 301-9.

117.    Kumar, A., et al., *Volumetric asymmetries in late-onset mood disorders: an attenuation of frontal asymmetry with depression severity.* Psychiatry Res, 2000. **100**(1): p. 41-7.

118.    Andreescu, C., et al., *Significance of brain MRI volumetric changes in late life depression: toxic stress or Dementia prodrome?* Neuropsychopharmacology (in press), 2008.

119.    Butters, M.A., et al., *The nature and determinants of neuropsychological functioning in late-life depression.* Arch Gen Psychiatry, 2004. **61**(6): p. 587-95.

120.    Taylor, W.D., et al., *Late-life depression and microstructural abnormalities in dorsolateral prefrontal cortex white matter.* Am J Psychiatry, 2004. **161**(7): p. 1293-6.

121.    Alexopoulos, G.S., et al., *Frontal white matter microstructure and treatment response of late-life depression: a preliminary study.* Am J Psychiatry, 2002. **159**(11): p. 1929-32.

122.    Salloway, S., et al., *The relationship of MRI subcortical hyperintensities to treatment response in a trial of sertraline in geriatric depressed outpatients.* Am J Geriatr Psychiatry, 2002. **10**(1): p. 107-11.

123.    Alexopoulos, G.S., et al., *'Vascular depression' hypothesis.* Arch Gen Psychiatry, 1997. **54**(10): p. 915-22.

124.    Krishnan, K.R., J.C. Hays, and D.G. Blazer, *MRI-defined vascular depression.* Am J Psychiatry, 1997. **154**(4): p. 497-501.

125.    Kumar, A., et al., *Regional cerebral glucose metabolism in late-life depression and Alzheimer disease: a preliminary positron emission tomography study.* Proc Natl Acad Sci U S A, 1993. **90**(15): p. 7019-23.

126.    Smith, G.S., et al., *The functional neuroanatomy of geriatric depression.* Int J Geriatr Psychiatry, 2009. **24**(8): p. 798-808.

127.    Raichle, M.E., et al., *A default mode of brain function.* Proc Natl Acad Sci U S A, 2001. **98**(2): p. 676-82.

128.    Damoiseaux, J.S., et al., *Reduced resting-state brain activity in the "default network" in normal aging.* Cereb Cortex, 2007.

129.    Greicius, M.D., et al., *Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI.* Proc Natl Acad Sci U S A, 2004. **101**(13): p. 4637-42.

130.    Greicius, M.D., et al., *Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus.* Biol Psychiatry, 2007. **62**(5): p. 429-37.

131.    Sheline, Y.I., et al., *The default mode network and self-referential processes in depression.* Proc Natl Acad Sci U S A, 2009. **106**(6): p. 1942-7.

132. Zhao, X.H., et al., *Altered default mode network activity in patient with anxiety disorders: an fMRI study.* Eur J Radiol, 2007. **63**(3): p. 373-8.

133. Alexopoulos, G.S., *Depression in the elderly.* Lancet, 2005. **365**(9475): p. 1961-70.

134. Aizenstein, H.J., Butters, M. A., Wu, M., Mazurkewicz, L. M., Stenger, V. A., Gianaros, P. J., Becker, J.T., Reynolds, C.F. III, Carter C.S. , *Altered Functioning of The Executive Control Circuit in Late-Life Depression: Episodic and Persistent Phenomena.* American Journal of Geriatric Psychiatry., In Press.

135. Reynolds, C.F., 3rd, et al., *Maintenance treatment of major depression in old age.* N Engl J Med, 2006. **354**(11): p. 1130-8.

136. First, M., et al., *Structured clinical interview for DSM-IV Axis I Disorders - Patient Edition (SCID-I/P). Version 2.0.* New York: New York State Psychiatric Institute., 1995.

137. Hamilton, M., *A rating scale for depression.* J Neurol Neurosurg Psychiatry, 1960. **23**: p. 56-62.

138. Folstein, M.F., S.E. Folstein, and P.R. McHugh, *"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.* J Psychiatr Res, 1975. **12**(3): p. 189-98.

139. Mattis, S., *Dementia Rating Scale-2: Professional Manual.* Psychological Assessment Resources: Odessa, FL, 2004.

140. Fair, D.A., et al., *A method for using blocked and event-related fMRI data to study "resting state" functional connectivity.* Neuroimage, 2007. **35**(1): p. 396-405.

141. Greicius, M.D., et al., *Functional connectivity in the resting brain: a network analysis of the default mode hypothesis.* Proc Natl Acad Sci U S A, 2003. **100**(1): p. 253-8.

142. Wu, M., et al., *A fully automated method for quantifying and localizing white matter hyperintensities on MR images.* Psychiatry Res, 2006. **148**(2-3): p. 133-42.

143. Holmes, C.J., et al., *Enhancement of MR images using registration for signal averaging.* J Comput Assist Tomogr, 1998. **22**(2): p. 324-333.

144. Cox, K., H.J. Aizenstein, and J. Fiez, *Striatal outcome processing in healthy aging.* CABIN  In Press.

145. Lowe, M.J., B.J. Mock, and J.A. Sorenson, *Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations.* Neuroimage, 1998. **7**(2): p. 119-32.

146. Mazoyer, B., et al., *Cortical networks for working memory and executive functions sustain the conscious resting state in man.* Brain Res Bull, 2001. **54**(3): p. 287-98.

147. Cox, R.W., *AFNI: software for analysis and visualization of functional magnetic resonance neuroimages.* Comput Biomed Res, 1996. **29**(3): p. 162-73.

148. Mayberg, H.S., et al., *Regional metabolic effects of fluoxetine in major depression: serial changes and relationship to clinical response.* Biol Psychiatry, 2000. **48**(8): p. 830-43.

149. Drevets, W.C., et al., *Subgenual prefrontal cortex abnormalities in mood disorders.* Nature, 1997. **386**(6627): p. 824-7.

150. Gusnard, D.A., et al., *Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function.* Proc Natl Acad Sci U S A, 2001. **98**(7): p. 4259-64.

151. D'Argembeau, A., et al., *Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking.* J Cogn Neurosci, 2007. **19**(6): p. 935-44.

152. Butters, M.A., et al., *Executive functioning, illness course, and relapse/recurrence in continuation and maintenance treatment of late-life depression: is there a relationship?* Am J Geriatr Psychiatry, 2004. **12**(4): p. 387-94.

153. Taragano, F.E., P. Bagnatti, and R.F. Allegri, *A double-blind, randomized clinical trial to assess the augmentation with nimodipine of antidepressant therapy in the treatment of "vascular depression".* International Psychogeriatrics, 2005. **17**(3): p. 487-498.

154. Fischl, B. and A.M. Dale, *Measuring the thickness of the human cerebral cortex from magnetic resonance images.* Proc Natl Acad Sci U S A, 2000. **97**(20): p. 11050-5.