The Effects of Amygdalar Size Normalization on Group Analysis in Late-Life Depression

R.J. Tamburo[1,*], J.T. Becker[1], G.J. Siegle[1], M.A. Butters[1], C.F. Reynolds III[1], H.J. Aizenstein[1]

Department of Psychiatry, University of Pittsburgh

Abstract

Structural MRI has been utilized in numerous ways to measure morphologic characteristics of subcortical brain regions. Volumetric analysis is frequently used to quantify the size of brain structures to ultimately compare size differences between individuals. In order to make such comparisons, inter-subject variability in brain and/or head size must be taken into consideration. A heterogeneous set of methods are commonly used to normalize regional volume by brain and/or head size yielding inconsistent findings making it difficult to interpret and compare results from published volumetric studies.

This study investigated the effect that various volume normalization methodologies might have on group analysis. Specifically, the amygdalae were the regions of interest in elderly, healthy and depressed individuals. Normalization methods investigated included spatial transformations, brain and head volume, and tissue volume techniques. Group analyses were conducted with independent t-tests by dividing amygdalar volumes by various volume measures, as well as with univariate analysis of covariance (ANCOVA) analyses by using amygdalar volumes as dependent variables and various volume measures as covariates. Repeated measures ANOVA was performed to assess the effect of each normalization procedure.

Results indicate that volumetric differences between groups varied based on the normalization method utilized, which may explain, in part, the dis-

[☆]No conflicts of interest exist for this study.

[*]Corresponding Author

Email addresses: rjtst21@pitt.edu (R.J. Tamburo), BeckerJT@upmc.edu (J.T. Becker), gsiegle@pitt.edu (G.J. Siegle), buttersma@upmc.edu (M.A. Butters), ReynoldsCF@upmc.edu (C.F. Reynolds III), aizen@pitt.edu (H.J. Aizenstein)

crepancy found in amygdalar volumetric studies. We believe the findings of this study are extensible to other brain regions and demographics, and thus, investigators should carefully consider the normalization methods utilized in volumetric studies to properly interpret the results and conclusions.

Keywords: Amygdala, morphometry, depression, structural MRI

1. Introduction

The development of medical imaging technology has given the medical community a powerful, non-invasive tool for viewing internal human anatomy. Medical imaging assists clinicians in making diagnoses and researchers in learning more about various pathologies. Computational methodologies have been developed to further assist clinicians and researchers in their endeavors by providing techniques for visualization, analysis, and quantification. For example, surgeons are able to plan and practice their procedures before entering the operating room, image rendering allows for three-dimensional visualization of anatomical structures, and segmentation allows for anatomical structures to be quantified via various measures like size (volumetry).

Neuroimaging researchers have made extensive use of structural magnetic resonance imaging (MRI) for studying subcortical brain structure morphology in the presence of various neurodegenerative diseases and psychological mood disorders. Late-life depression (LLD) is of particular clinical interest because it is the most common psychiatric illness affecting the elderly population (older than 65 years of age) in the United States. The National Institutes of Health estimates that 5.7% of the elderly population suffer from major depression and another 14.2% exhibit clinically relevant depressive symptoms (Health, 2007). Individuals suffering from LLD exhibit common depressive symptoms such as prolonged changes in mood and behavior. Additionally, LLD frequently coexists with other medical illnesses such as cognitive impairment Butters et al. (2008), and negatively impacts quality of life.

The Health and Retirement Study (HRS) reports that the prevalence of clinically relevant depressive symptoms in the civilian noninstitutionalized elderly population from 1998 to 2004 has gradually declined from 11.9% to 11.0% in men and 18.6% to 16.8% in women (15.9% to 14.4% overall). However, the etiology of LLD is not fully understood and thus remains a major health problem because it continues to be underrecognized, misdiagnosed,

and untreated. The individual and societal burdens associated with LLD are expected to amplify as the aging population increases. Projections by the Population Division of the U.S. Census Bureau (NP2008-T12), estimate that from 2010 to 2030 the elderly population will increase from 13% to 19% of the U.S. population nearly doubling in size Aging (2007). The expected increased prevalence of LLD warrants energized research to meet the anticipated demand for new and effective diagnosis and treatment.

Kluver, et al, Kluver and Bucy (1939), in their pioneering amygdalar lesion research, demonstrated the role of the amygdalae's effect on mood. Since that time, the amygdalae have been a focus of depression research due to their key role in integrating emotional meaning with perception and experience. Observations of the amygdalae's functional role in mood disorders have been quantified and verified in electrophysiological (e.g., electroencephalography and magnetoencephalography) and neuroimaging (e.g., functional magnetic resonance imaging and positron emission tomography) studies. Some mood disorder researchers have used MRI to investigate the anatomical basis for these functional observations by measuring changes in amygdalar size. Results from these studies, however, have been variable and inconclusive.

For example, there are reports of 1) increased volume for both amygdalae Frodl et al. (2002); Lange and Irle (2004) and the right amygdala Bremner et al. (2000) with major depression, bipolar disorder Altshuler et al. (1998), first-episode nonschizophrenic psychosis Velakoulis et al. (2006), and generalized anxiety disorder Bellis et al. (2000); 2) decreased amygdalar volume in depression Rosso et al. (2005), mild dementia Hensel et al. (2005), depression with memory problems Gunten et al. (2000), depression with psychosis Keller et al. (2008), a unilateral volume decrease in the left amygdala in mild dementia Hensel et al. (2005) and depression with memory problems Gunten et al. (2000), and a unilateral volume decrease in the right amygdala with major depression Xia et al. (2004); Hastings et al. (2004); and 3) no change in amygdalar volume in major depression Bremner et al. (2000); Mervaala et al. (2000); Munn et al. (2007); Hastings et al. (2004); Tamburo et al. (2008), questionable dementia Hensel et al. (2005), recurrent major depression Sheline et al. (1998), and depression without psychosis Keller et al. (2008). A compilation of findings from total amygdalar volumetric studies is summarized in table 3 to document the variability of these findings along with the range of MRI parameters, amygdalar segmentation protocols, number of subjects, age of subjects, and normalization methods used for each study.

The mixed results obtained in these volumetric studies may be attributable to a variety of factors. In a prior study Tamburo et al. (2008), we explored the possibility that reported volumetric discrepancies may be due to an inherent limitation in considering a gross measure of size rather than shape. In that study, significant atrophy was localized to the basolateral nucleus in persons with LLD despite insignificant volumetric findings. Results from volumetric studies are dependent on clinical and demographic variables. Clinical and demographic variables sometimes, but not always, considered include psychiatric disorder (column 2, table 3), gender (column 5, table 3), age (column 6, table 3), as well as number of depressive episodes, depression history, substance abuse, childhood trauma, etc.

The common analytic process used for neuroimaging volumetric studies warrants investigation as well. The following procedure is typically followed: 1) Segmentation to extract the structure(s) of interest, 2) Structure size normalization to account for subject variability, and 3) Statistical analysis to compare structure volumetrics between groups. Each of these steps may be carried out in a variety of ways adding to the difficulty of directly comparing results between published studies.

Perhaps the most critical factor affecting volumetric measurements is the accuracy of segmenting the amygdalae. Segmentation accuracy is affected by the quality of the acquired image, e.g., image resolution (column 3, table 3) and the anatomic definition of the amygdala. The inclusion or exclusion of various nucleic subdivisions is still debated amongst neuroscientists and some boundaries of the amygdala are not distinctly visible in MRI images. Consequently, researchers have devised and employed a number of delineation protocols to segment the amgydalae as enumerated in column 4 of table 3.

To account for intrinsic anatomic variability between individuals, neuroimaging researchers sometimes attempt to remove inter-subject variability by normalizing volumetric measurements. Similar to the range of segmentation methods used, a large number of techniques have been employed for volumetric normalization (column 7, table 3). Common normalization techniques utilize head or brain volume measures to divide structure volume by or include as covariates in statistical analyses. We hypothesize that different normalization techniques produce inconsistent results, and consequently affect the interpretation of results and conclusions between published studies. This study explores the effects of various strategies for amygdalar size normalization on volumetric analysis in LLD. Although the brain region of interest is the amgydalae and the clinical pathology is LLD, findings should

be considered for analysis of other brain structures and pathologies.

Table 1: Summary of published total amygdalae volumetric results in mood disorder research. [1]MRI Prop: MRI properties reported here are MRI magnet strength (Tesla, T) followed by slice thickness in millimeters (mm). [2]Seg Meth: Reference to the segmentation protocol used to delineate the amygdalae. [3]N: The total number of subjects included in the study counted by gender. [4]Age: Mean age (standard deviation) for mood disorder subjects followed by control subjects. [5]Findings: Total amygdalar volume differences are reported as mood disorder relative to control. The '+' denotes an increase in size, − denotes a decrease in size, and Ø denotes no change in size. B denotes a bilateral difference and L/R denotes a unilateral difference for the left/right amygdala. Significance is taken at $p < 0.05$ level. $\star$ indicates that no gender differences were observed and † indicates that gender differences were not assessed. [6]Norm Method: Normalization method followed by statistical test. None = no normalization method. TICV = total intra-cranial volume. TBV = total brain volume. MBAR = mean brain area ratio. SN = spatial normalization.

Study	**Disorder**	**MRI Prop**[1]	**Seg Meth**[2]	**N**[3]	**Age**[4]	**Findings**[5]	**Norm Method**[6]
Frodl et al. (2002)	Major	1.5 T	Convit et al. (1999)	34 F	40.3 (12.6)	+ L*	None (t-test)
	Depression	1.5 mm		26 M	40.6 (12.5)	+ B*	TBV without stem (t-test)
						+ B*	TICV (ANCOVA)
Lange and Irle (2004)	Major	1.5 T	Pruessner et al. (2000)	34 F	34 (10)	+ B†	TBV (ANCOVA)
	Depression	1.3 mm		0 M	32 (6)		
Xia et al. (2004)	Major	1.5 T	Xia et al. (2004)	18 F	39.5 (12)	− R†	None (t-test)
	Depression	1.2 mm		17 M	35.4 (8.9)		
Hastings et al. (2004)	Major	1.5T	Sheline et al. (1998)	20 F	38.9 (11.4)	Ø B (M)	TBV (ANOVA)
	Depression	1.5 mm		16 M	34.8 (13.6)	− R (F)	
Bremner et al. (2000)	Major	1.5 T	Bremner et al. (1995)	12 F	43 (8)	+ R†	None (ANOVA)
	Depression	3 mm		20 M	45 (10)	Ø B†	TBV (ANOVA)
Mervaala et al. (2000)	Major	1.5 T	Watson et al. (1992)	29 F	42.2 (12.2)	Ø B†	None (t-test)
	Depression	3 mm		22 M	42.1 (14.6)	Ø B†	MBAR (t-test)
Munn et al. (2007)	Major	1.5 T	Munn et al. (2007)	44 F	20.54 (1.75)	Ø B†	None (t-test)
	Depression	1 mm		0 M	20.73 (2.07)		
Keller et al. (2008)	Depression	3 T	Kates et al. (1997)	23 F	36.6 (11.9)	Ø B†	None (ANCOVA)
	without Psychosis	1.5 mm		18 M	32.2 (11.5)		

2. Methods

2.1. Subject Characteristics

To investigate amygdalar volumetrics in LLD, MRI data were acquired from LLD diagnosed patients and healthy elderly controls. A total of 35 subjects were included in this study and they all gave informed consent prior to participation in this study. All but one of the participating subjects were right-handed. Each subject received a Structured Clinical Interview for DSM-IV (SCID-IV) evaluation, which were reviewed in a diagnostic consensus conference. Exclusion criteria included all Axis I psychiatric disorders except for major depressive disorder and anxiety disorders (for subjects in the LLD group). Subjects with co-morbid anxiety disorders were included due to the high prevalence (48%) of anxiety disorders in subjects with LLD Beekman et al. (2000). Cognitive status was assessed using the Mattis Dementia Scale (MDS). See table 5 for MDS, Hamilton, and Mini-Mental State Exam scores, and other demographic information. MDS, Hamilton, and MMSE scores were not available for 5 of the control group subjects (2 male and 3 female). Subjects were also excluded for a history of stroke or significant head injury, dementia, Alzheimer's disease, Parkinson's disease, or Huntington's disease.

The LLD subject group consisted of 16 elderly individuals; 9 males and 7 females with an average age of 67.9 $\pm$ 5.5 years (mean $\pm$ standard deviation). Each of these individuals diagnosed with LLD had a history of being medicated with an antidepressant. The control group consisted of 19 healthy elderly subjects; 11 males and 8 females with an average age of 67.2 $\pm$ 7.1 years. These control subjects presented with no clinical symptoms of LLD and did not meet any of the exclusion criteria.

2.2. MRI Data Acquisition

Each subject in the study had their head scanned using the same MRI scanning protocol, which was approved by the internal review board at the University of Pittsburgh. High resolution, 3D data were acquired on a 1.5 Tesla Signa Scanner (General Electrics Medial Systems, Milwaukee, WI) full body scanner. A spoiled GRASS imaging sequence was used with the following acquisition parameters: TR/TE = 5/25 ms, flip angle = 40°, and FOV = 24 $\times$ 18 cm. Images were acquired with the subject in the prone position and had a resolution of 0.9375 $\times$ 0.9375 mm in the axial plane and 1.5 mm in the inter-slice dimension.

2.3. Amygdalae Segmentation

For our volumetric analyses, both left and right amygdala were manually segmented from each subject's MRI scan. All images were initially aligned to be in-plane with the AC-PC (anterior commisure-posterior commisure); this was done using the rigid body, landmark-based transformation routine available in the Automated Functional NeuroImaging Cox (1996) software suite.

Following AC-PC alignment, a trained person manually delineated the boundaries of each amygdala following a protocol in which adequate intra- and inter-rater reliability has been previously established Siegle et al. (2002) (posterior boundary: the alveus of the hippocampus; anterior boundary: 2 mm from the temporal horn of the lateral ventrical; superior boundary: ventral horn of the sub-arachnoid space (SS); inferior boundary: most dorsal finger of the white matter tract under the horn of the SS; lateral boundary: 2 mm from the surrounding white matter; mesial: 2 mm from the SS). The tracer performing the manual segmentations was blind to the group status for each subject.

Shown in Figure 3 is an example amygdala segmentation rendered as a 3D surface model and overlaid on cross-sections of the corresponding MRI image. Segmentations of each subject's amygdalae were saved as a binary image and bisected into separate images such that the left and right amygdala could be analyzed independently. Amygdalar volumes were measured by counting the number of voxels within the delineated region and multiplying by the voxel resolution.

2.4. Image Processing

Multiple strategies for normalization were utilized, with each normalization strategy requiring a specific image processing pipeline. Consequently, a variety of image analysis techniques and software tools were used. The aim of this study is comparison of different normalization approaches. Although, the accuracy of the segmentation methods used in the different normalization methods was not a particular focus of this study, we chose parameters for segmentation which the literature has suggested provide accurate assessments. The methods are briefly discussed below.

2.4.1. Brain Segmentation

When processing and analyzing MRI brain images, it is beneficial to remove non-brain voxels from the image. In this study, segmentation of the

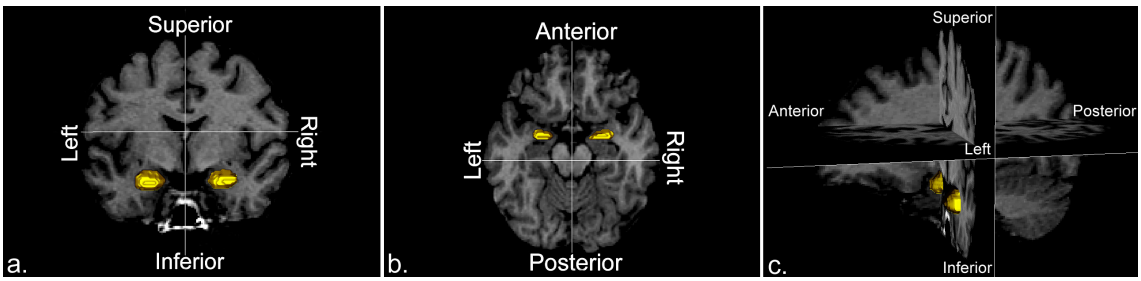

Figure 1: a. Coronal, b. axial, and c. orthogonal cross-sections of a subject's MRI image are shown with 3D renderings of their segmented amygdalae. Visualization performed with in-house software designed with the Insight Toolkit (http://www.itk.org), Fast Light Toolkit (http://www.fltk.org), and Visualization Toolkit (http://www.vtk.org).

brain, sometimes referred to as skull stripping, is the first step in each processing pipeline. By removing from the image as many voxels not belonging to white matter (WM), gray matter (GM), or cerebrospinal fluid (CSF) of the brain, computational methods and resulting measures are more reliable. The FMRIB Analysis Group at the University of Oxford has developed an algorithm for automated brain segmentation and has made the software implementation of the algorithm called the Brain Extraction Tool (BET) freely available for non-commercial use Smith (2002). BET was used in this study because it is automated, widely used in neuroimaging, fast, requires no preprocessing, and performs reasonably well.

BET uses a deformable model that evolves to the brain's surface by the application of a set of locally adaptive model forces. The deformable model begins at the estimated brain center as a sphere. To calculate the center of the brain, the image is thresholded to exclude voxels with intensities in the bottom and top 2% of the cumulative histogram. The center position of the brain is then estimated by computing the center of gravity (COG) of the image, excluding the background, as the standard intensity-weighted sum of voxel locations. The size of the sphere is estimated by approximating the radius of the brain from non-background voxels. Once the model is initialized, it iteratively deforms until the outer skull is reached resulting in a whole brain segmentation.

BET has moderate success when executed with its default set of parameters. However, BET tends to include unwanted voxels in the final segmenta-

tion if the brain COG calculation is inaccurate due to many non-brain voxels contained within the image, e.g., if the scanning field of view includes too much of the neck. BET does offer a number of parameters to control the performance of the segmentation algorithm including one for manually setting the brain COG (BET option: $-c < x\ y\ z >$, where x, y, and z are the voxel coordinates for the COG).

Fagiolo, et al. Fagiolo et al. (2008) have developed a recursive method for deducing an optimal brain COG. Their method initially uses the COG estimate and brain segmentation from BET with all default parameter settings. In successive iterations, the COG is recalculated from the resulting brain segmentation in the prior iteration. The iteration ceases when the Euclidean distance between COG estimates is half the diagonal of a voxel. Fagiolo, et al. report that in just a few iterations, this method dramatically increases the performance of BET for automated brain segmentation.

To illustrate the performance of these brain segmentation methods, a subject image was automatically segmented with BET using the default parameters and with BET using the recursive method. The resulting segmentations are shown in Figure 4. Figure 4a shows the automatic segmentation result using BET with default parameters and Figure 4b shows the automatic segmentation result with the recursive BET algorithm after 4 iterations. Note the neck region near the brain stem.

In this study, brains were segmented with BET using both the default parameters (BET-default) and the recursive method (BET-recursive) by Fagiolo, et al. The recursive BET method was implemented with the Insight Toolkit Yoo (2004). Recent versions of the BET software application now include a recursive option (BET option: $-R$).

2.4.2. Brain and Head Segmentation

Neuroimaging researchers sometimes normalize brain structure volume by gray matter, white matter, and cerebrospinal fluid volume. In order to quantify these regions they must first be segmented. There are a large number of computational methods for automated segmentation each with varying degrees of accuracy and computational speed. Some of these methods are freely available in neuroimaging software packages. In this study, we used a popular software package called FMRIB's Automated Segmentation Tool (FAST) Zhang et al. (2001).

Segmentation with FAST (version 3.53) was performed with default parameter settings (FAST options: $-t\ 1\ -c\ 3$) on brain images segmented with

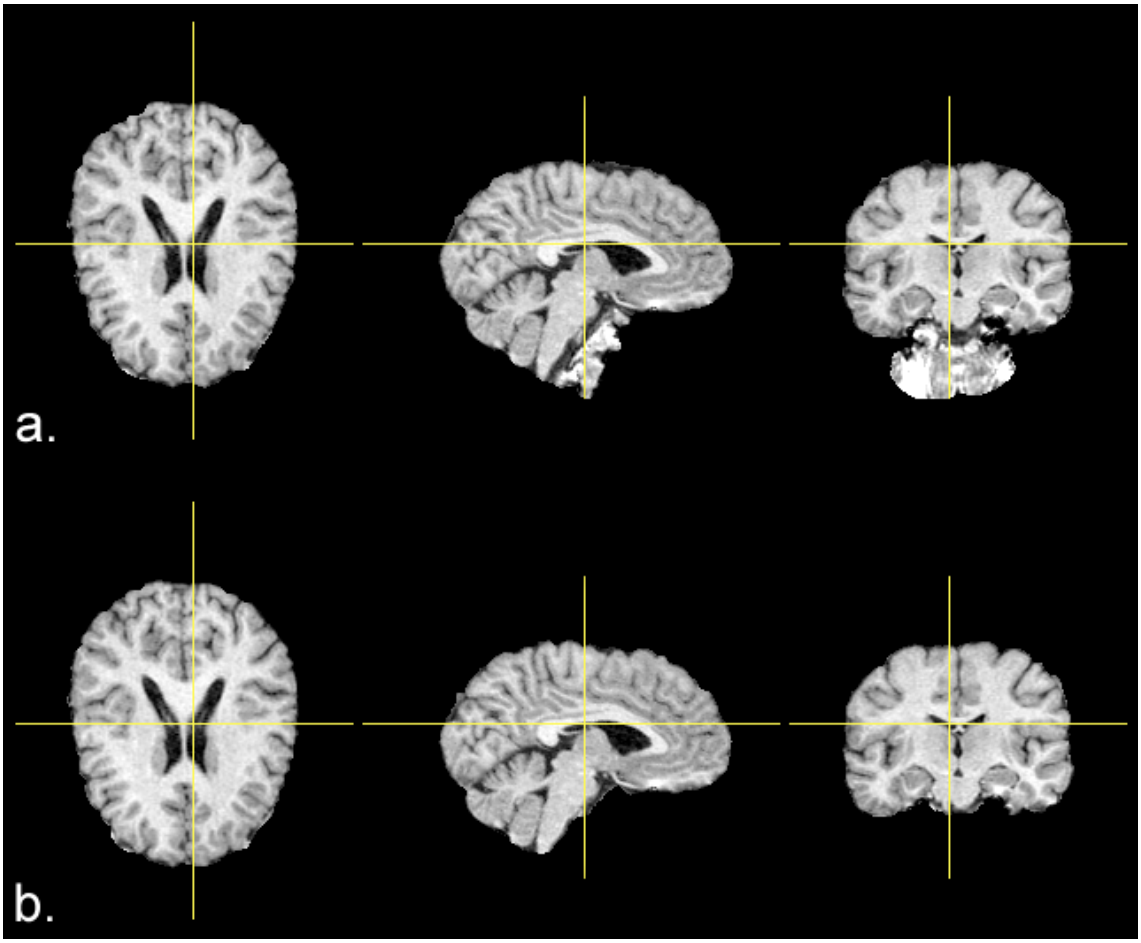

Figure 2: a. Image automatically segmented using BET with the default parameter set and b. the same image automatically segmented using BET recursively. Visualization performed with ImageJ (http://rsbweb.nih.gov/ij).

BET-default and BET-recursive. Volumes were calculated from the probabilistic segmentation images (FAST option: $-op$) by summing classification probabilities and multiplying by voxel volume.

In addition to normalizing amygdalar volumes by tissue segmentation volumes, amgydalar volumes were also normalized by measures of brain and head size. Total brain volume (TBV) was calculated as the sum of GM, WM, and CSF. Total brain volume including CSF volume within the entire skull is often referred to as total intracranial volume (TICV). Some neuroimaging researchers, particularly those that study elderly mood disorders, have noted that TICV may be a more suitable method for normalization because it is invariant to brain atrophy like TBV since TICV measures the volume contained within the skull, which does not normally change throughout adulthood.

Brain segmentation software usually includes the CSF within the surface of the brain, but not necessarily the CSF between the skull and the brain. In order to measure TICV, BET was used with advanced option $-A$ to extract additional skull and scalp surfaces Jenkinson et al. (2005). A volume measure qualitatively equivalent to TICV was calculated as the volume contained within BET's 'inner skull' segmentation. Also segmented with BET was the 'outer skull' surface, the enclosed volume of which is referred to as the total cranial volume (TCV) and includes the TICV and skull volume.

Also calculated was the total head volume (THV) volume, which is the volume contained within BET's 'outer skin' surface segmentation, i.e., the person's visible head. The outer and inner skull volume measures are susceptible to inconsistent imaging field of view across subjects. Since the images were ACPC aligned, the volumes were made invariant to image field of view by excluding segmentation voxels below the inferior surface of the cerebellum. See Figure 5 for an example of these segmentations. The accuracy of BET for approximating TBV, TICV, TCV, and THV was qualitatively assessed to be reasonable for the purposes of this paper.

2.4.3. Spatial Normalization

In addition to normalizing amygdalar volume with a variety of volumetric measures, spatial normalization was also explored. Subject images were registered to a study specific template with 7- and 12- degrees of freedom (DOF) transforms. The 7-DOF transform accounts for translation and rotation in three dimensions and isotropic scaling. The 12-DOF transform in three dimensions accounts for translation and rotation, and independent scaling in each of 3 cardinal directions. After image registration was per-

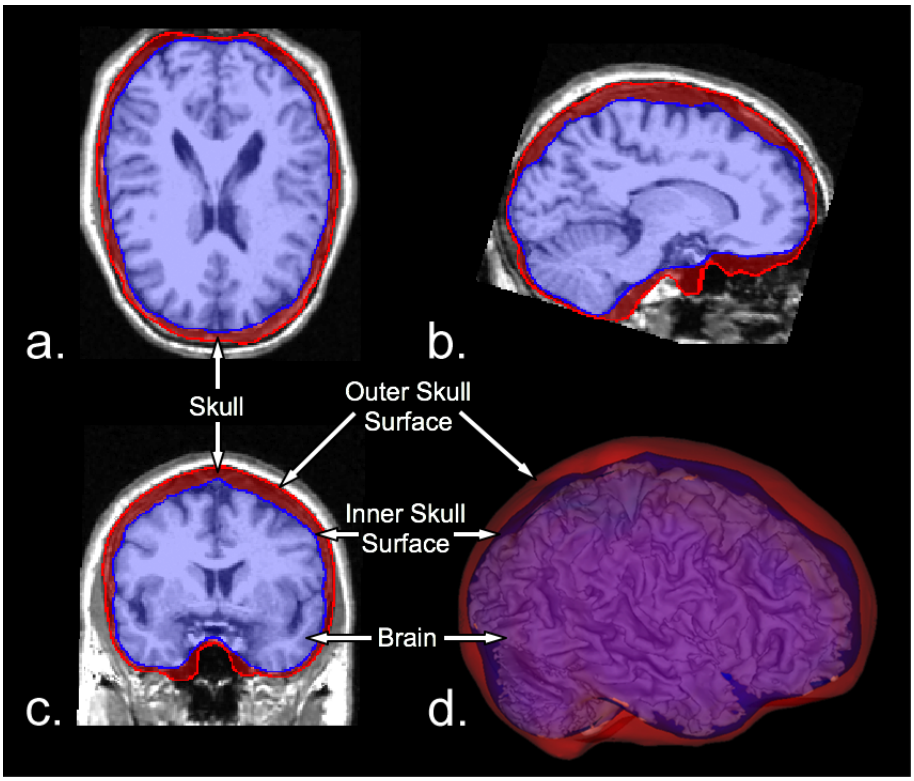

Figure 3: Example of BET brain and surface segmentations displayed on a. Axial, b. Sagittal and c. Coronal cross-sections of the image, and a d. three-dimensional rendering. The red line indicates the outer skull boundary and the blue line indicates the inner skull boundary. The skull is colored in red and the brain is colored in blue. Visualization performed with FSLView (http://www.fmrib.ox.ac.uk/fsl/fslview).

formed, the transformation matrix was applied to amygdalar segmentations. Interpolation was performed with the nearest neighbor technique since the segmentations were binary images. Amygdalar volumes were then measured as the sum of voxels multiplied by voxel volume.

The study specific template was generated using subject data. To avoid registration bias an equal number of control and LLD subjects were used to create the template. The template was generated using the brain segmentation images from BET-default. All of the images were aligned to an arbitrarily chosen reference image with a rigid body transform conducted with FSL's FLIRT (*flirt -cost normmi -dof 6*) Jenkinson et al. (2002) application. Once aligned, images were averaged together to produce the template using FSL's utility suite (fslmerge (*fslmerge -t merged_output *inputImages** and fslmaths *fslmaths merged_output -Tmean template*). See Figure 6 to view the study specific template.

2.5. Normalization Pipeline

The collection of the aforementioned methods to investigate amygdalar volume normalization includes 8 volume measures (GM, WM, GM + WM, GM + CSF, TBV, TICV, THV, and TCV) without and with 7-DOF and 12-DOF spatial transformation, and 2 methods of brain segmentation. A schematic of these pipelines are shown in Figure 7.

2.6. Statistical Analysis

Right and left amygdalar volumes were compared between groups with statistical analyses that prevail in the current literature. The first method uses two-sample t-tests with unequal variance and sample size (Microsoft Excel 2004, Seattle, Washington; *ttest(Data1, Data2, 2, 3)*). Amygdalar volumes were divided by the various normalization volumes. Data were grouped by diagnosis (control versus depressed) with and without gender separation, and compared for group analysis. P-values are reported.

For the second method, right and left amygdalar volumes were subjected to univariate analysis of covariance (ANCOVA) with SPSS 17.0 (*Analyze* $\rightarrow$ *General Linear Model* $\rightarrow$ *Univariate*) to test for group significance in amygdalar volume while adjusting for various normalizing volumes. Amygdalar volumes were the dependent variables. Group and gender were used as fixed factors and the various volume measures were used as covariates. Separate analyses were conducted for each normalization volume since the volumes are highly correlated with each other. When adjusting for $GM + WM$ and

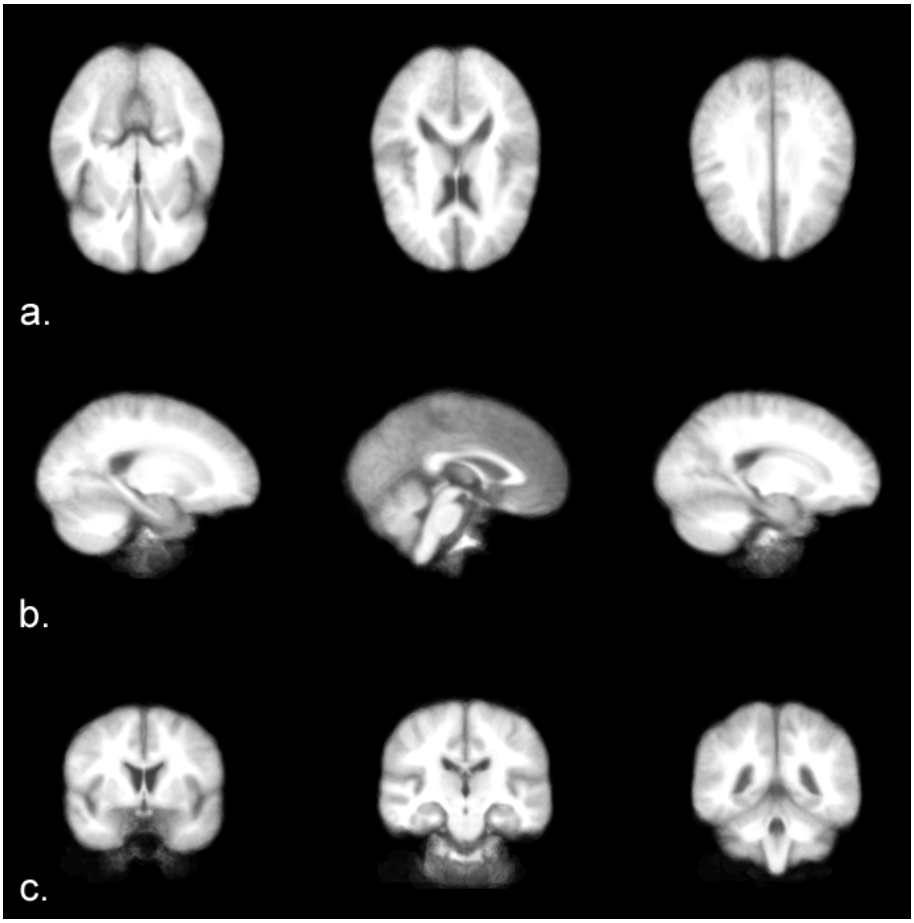

Figure 4: a. Axial, b. Sagittal, and c. Coronal cross-sections of the study-specific template. Images generated with FSL (http://www.fmrib.ox.ac.uk/fsl), i.e., *slicesdir 'imglob *'*.

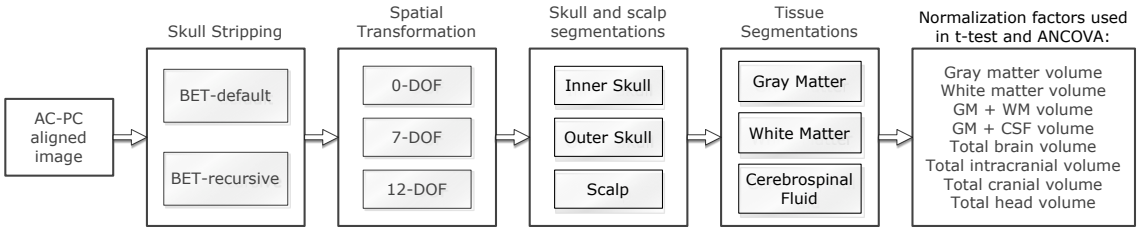

Figure 5: Schematic representation of the pipeline for amygdalar volume normalization.

$GM + CSF$, both volumetric measures were used as covariates. Reported are p-values for group and group by gender interaction effects.

For the third method, a repeated measures ANOVA was conducted (SPSS 17.0 (*Analyze* $\rightarrow$ *General Linear Model* $\rightarrow$ *Repeated Measures*) to test for significant effects of various within- and between-subject factors on amygdalar volume. The dependent variable was normalized amygdalar volume calculated as a ratio of amygdalar volume to TBV, TICV, TCV, and THV. Four within-subjects factors were investigated with two levels indicating side (right and left), three levels indicating number of degrees of transform (0, 7, and 12), four levels indicating normalization method (TBV, TICV, TCV, and THV), and two levels indicating stripping method (bet-default and bet-recursive). Group (control and depression) was used as a between-subjects factor.

Group means were compared to determine which group had the significantly larger/smaller means. For all statistical tests, a two-tailed p-value of 0.05 was chosen as the threshold for statistical significance.

3. Results and Discussion

3.1. Demographic and Clinical

Demographic and clinical data were analyzed with a two-tailed t-test to test for differences while grouped, separately, by diagnosis and gender. Statistical results are shown in Table 5. MMSE, Hamilton, and MDS scores were not available for 5 control group subjects (2 male and 3 female). Statistical results with $p < 0.05$ are considered significant and are shown in bold.

Results indicate no statistically significant group differences by age, education, or MMSE score. Hamilton scores were significantly greater in the LLD group with data pooled and separated by gender. MDS scores were significantly greater for the control group only with data pooled by gender.

3.2. Brain Segmentation and Spatial Transforms

The effect of brain segmentation method and spatial transforms on amygdalar, brain/head, and tissue volumetry was assessed by performing group-wise analysis with absolute volume measures. These results are tabulated in Supplementary Tables 1, 2, 3, 4, and 5. The right amygdala for LLD (combined and male) were consistently smaller for each trial, and the left amygdala was significantly smaller for LLD male subjects in all cases except for the trial with BET-default and 7-DOF. TBV and TCV were the only volumes

Table 2: Demographic and clinical information, and group analyses results. Data presented as mean (standard deviation), and p-values. Significance at $p < 0.05$ is shown in bold. *Age statistics calculated for all 35 subjects in study. **Education and MMSE, Hamilton, and MDS scores not available for 5 control group subjects.

Group (N)	**Age* (years)**	**Education** (years)**	**MMSE****	**Hamilton****	**MDS****
Control (14)	67.2 (7.1)	15.8 (2.7)	29.1 (0.86)	1.6 (1.6)	141 (2.4)
LLD (16)	67.9 (5.5)	14.4 (2.8)	28.4 (3.1)	14.4 (7.9)	138.2 (4.3)
p-value:	0.74	0.17	0.35	**9.2×10^{-6}**	**0.037**
Male Control (9)	68.2 (8.0)	16.0 (2.4)	29.0 (1.0)	1.9 (1.6)	140.4 (2.8)
Male LLD (9)	68.6 (6.0)	15.3 (2.4)	29.3 (1.3)	10.8 (8.0)	138.2 (4.5)
p-value:	0.93	0.56	0.56	**0.01**	0.23
Female Control (5)	65.8 (6.0)	15.6 (3.6)	29.4 (0.5)	1.2 (1.6)	142 (1.2)
Female LLD (7)	67.1 (5.2)	13.3 (3.1)	27.1 (4.3)	19.1 (5.1)	138.3 (4.3)
p-value:	0.64	0.28	0.22	**3.4×10^{-5}**	0.06

that exhibited group differences. The TBV of LLD females was greater than female controls with BET-default, 7-DOF. TCV was significantly greater in the LLD group with gender combined for both methods of brain segmentation and spatial transforms. Additionally, BET-default resulted in this group significance with both spatial transforms when BET-recursive did not. There were no significant group differences for any of the measured tissue volumes.

3.3. Amygdalar Normalization with t-tests

The results of right and left amygdalar volume normalization with t-tests are summarized in Supplementary Tables 6 and 7. Figures 8 and 9 illustrate, with gender combined, group analysis results for right and left amygdalar volume, respectively. Regardless of brain segmentation method and spatial transform, all normalization volumes resulted in significant group differences (LLD < Controls) for right amygdalar volume. Group significance (LLD < Controls) was variable for left amygdalar volume by brain segmentation method, normalization volume, and spatial transform. This finding could lead one to draw the conclusion that subjects with LLD have either a unilateral (right) or bilateral decrease in volume as compared with control subjects.

When subjects were separated by diagnosis, as well as gender, males exhibited a pattern similar to when subjects were pooled only by diagnosis. For females however, the right amygdalar volume was significant (LLD <

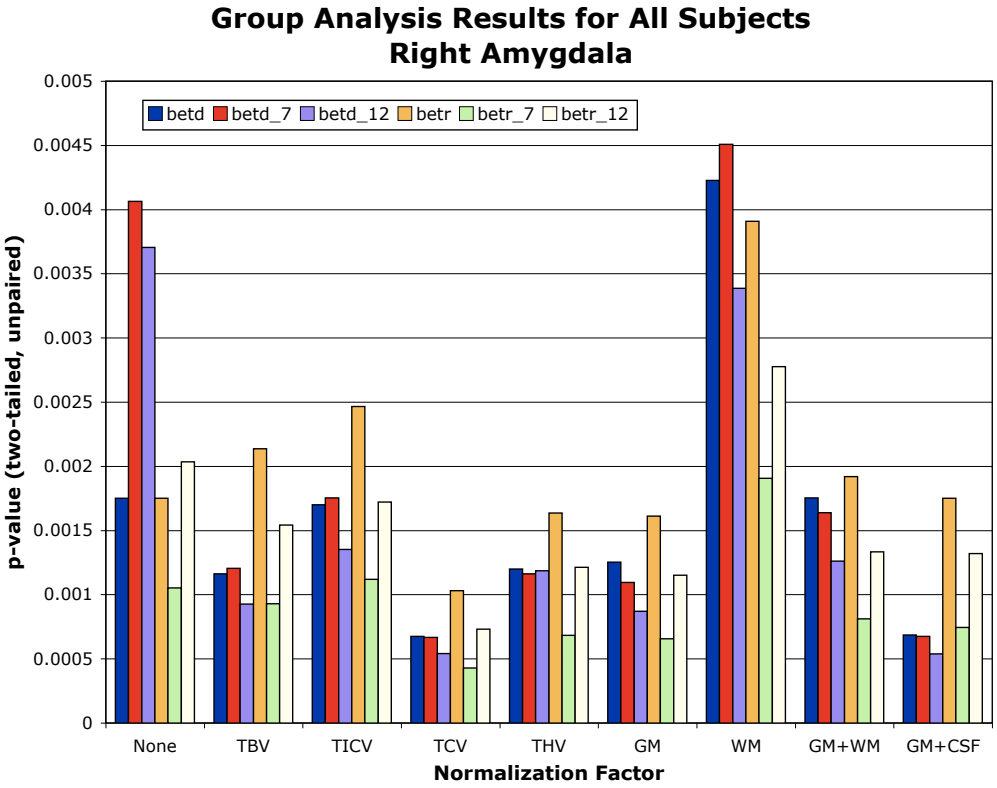

Figure 6: Group analysis results with all subjects combined for right amygdalar volume corrected by various normalization factors.

Controls) with a handful of normalization methods, and no left amygdalar volumes were significant at $p < 0.05$. The various normalization strategies, affected males differently than females yielding different group analysis results. With males, there were either unilateral (right) or bilateral differences (LLD < Controls). Group analysis results with normalized amygdalar volumes for females could be interpreted as not, unilaterally (right), or bilaterally significant smaller in LLD as compared to Controls.

The results obtained with subjects grouped by gender confirm that gender is a significant variable in amygdalar volumetry in LLD.

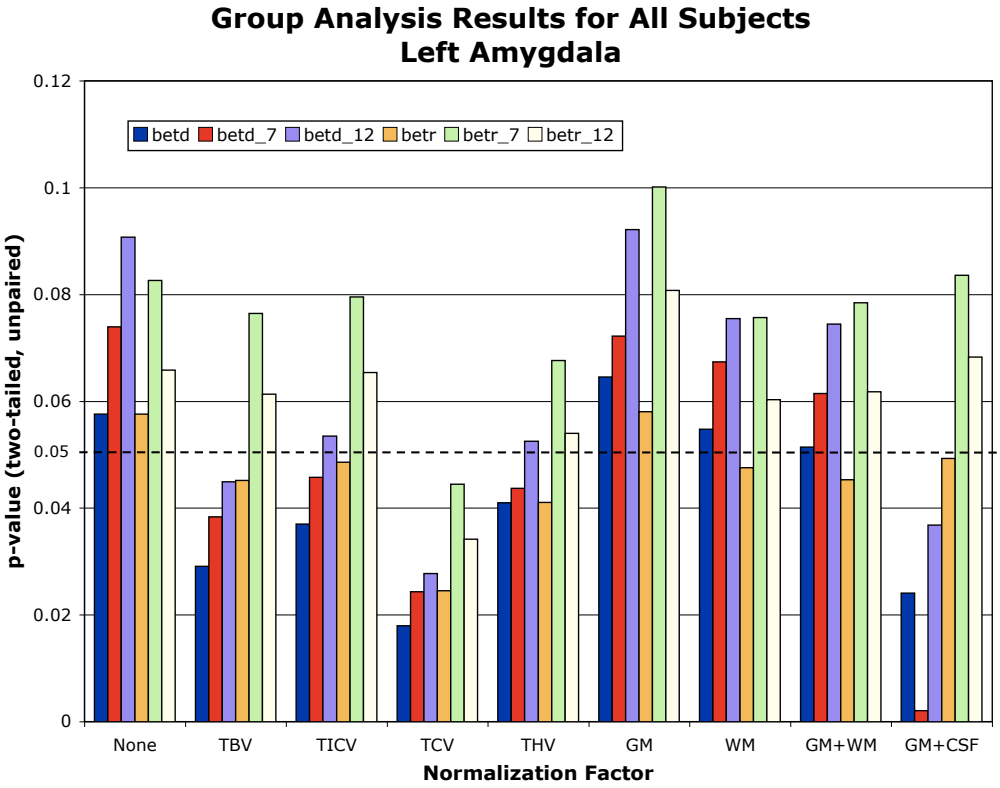

Figure 7: Group analysis results with all subjects combined for left amygdalar volume corrected by various normalization factors.

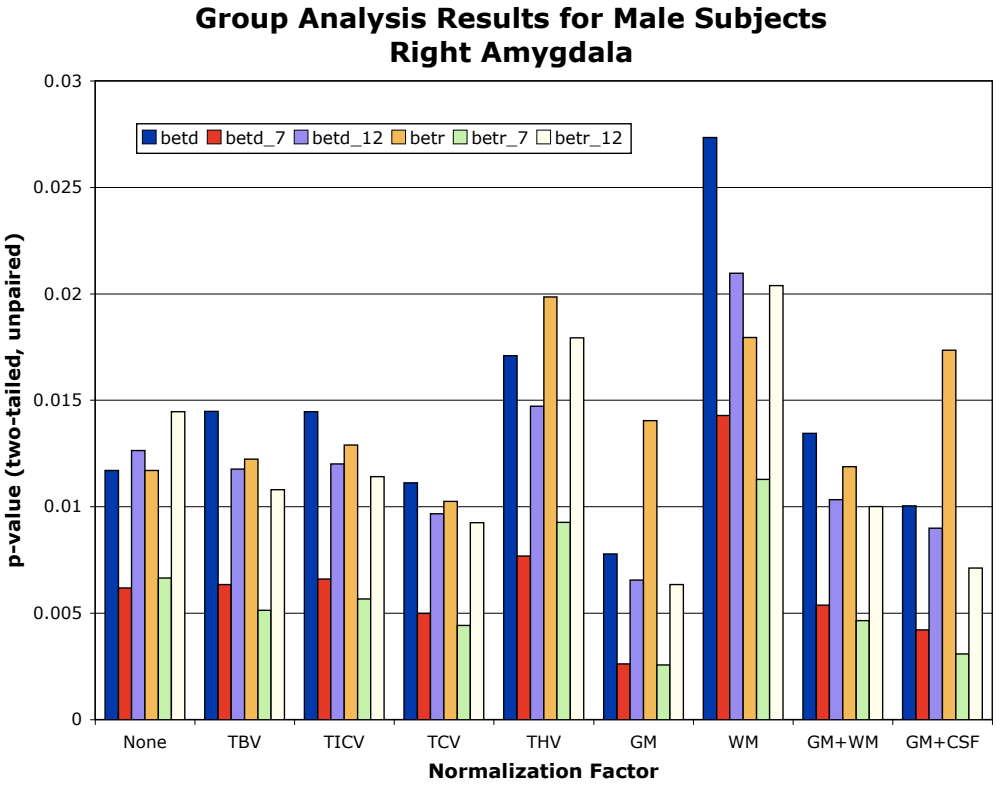

Figure 8: Group analysis results with male subjects for right amygdalar volume corrected by various normalization factors.

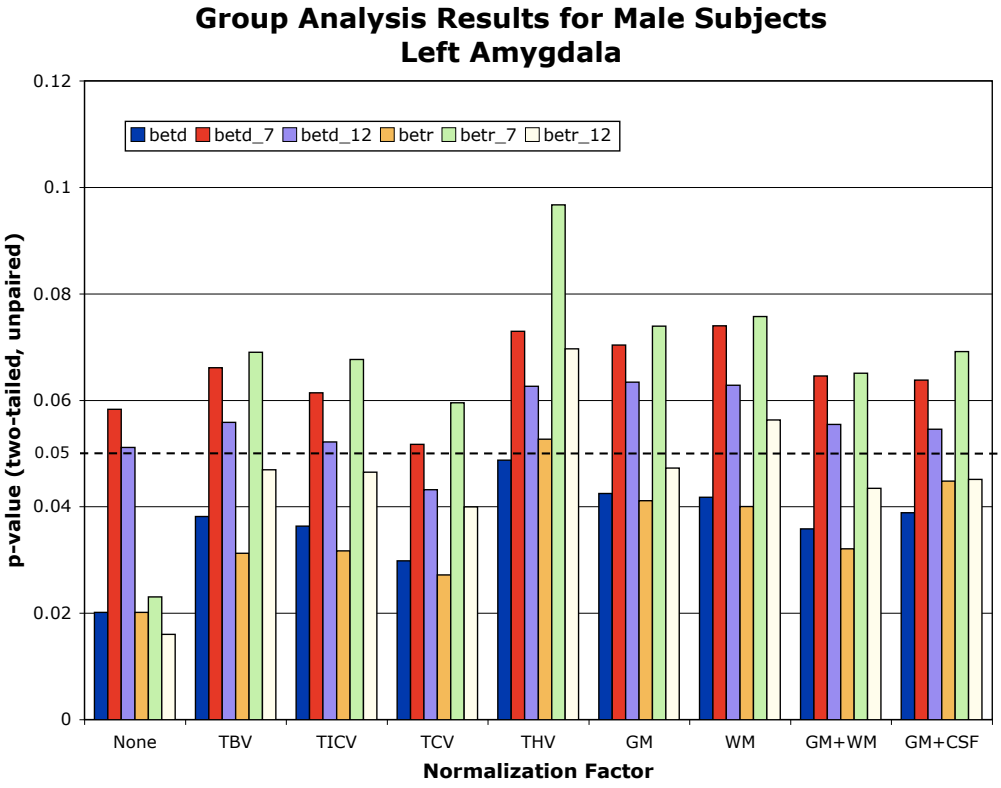

Figure 9: Group analysis results with male subjects for left amygdalar volume corrected by various normalization factors.

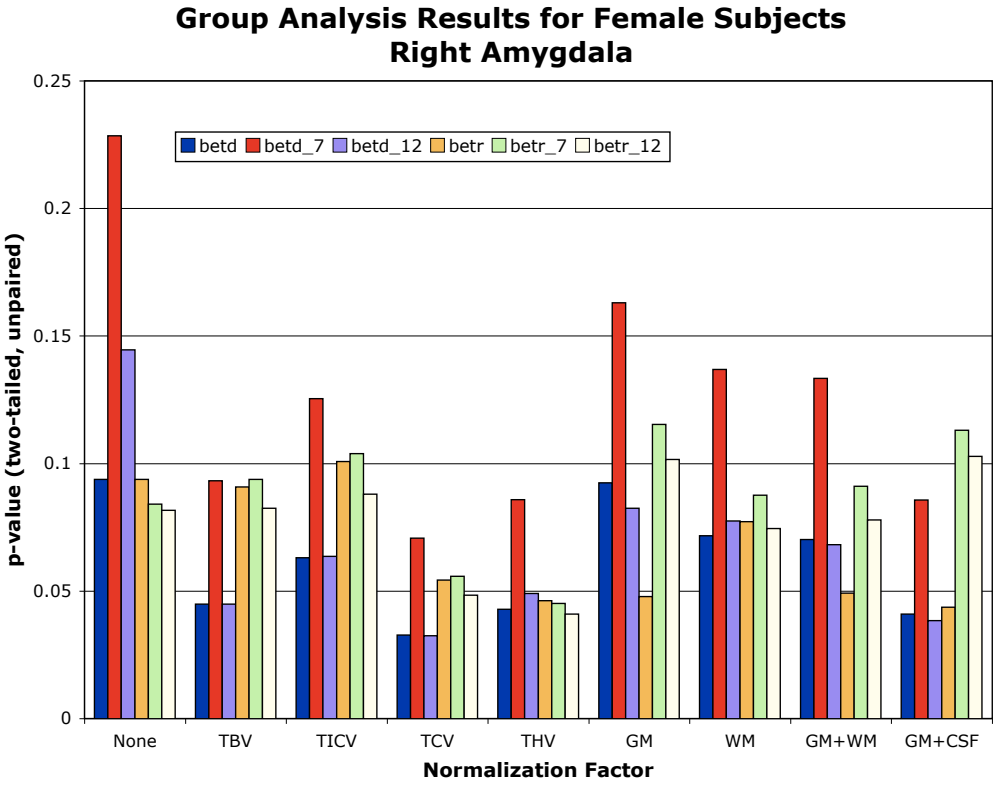

Figure 10: Group analysis results with female subjects for right amygdalar volume corrected by various normalization factors.

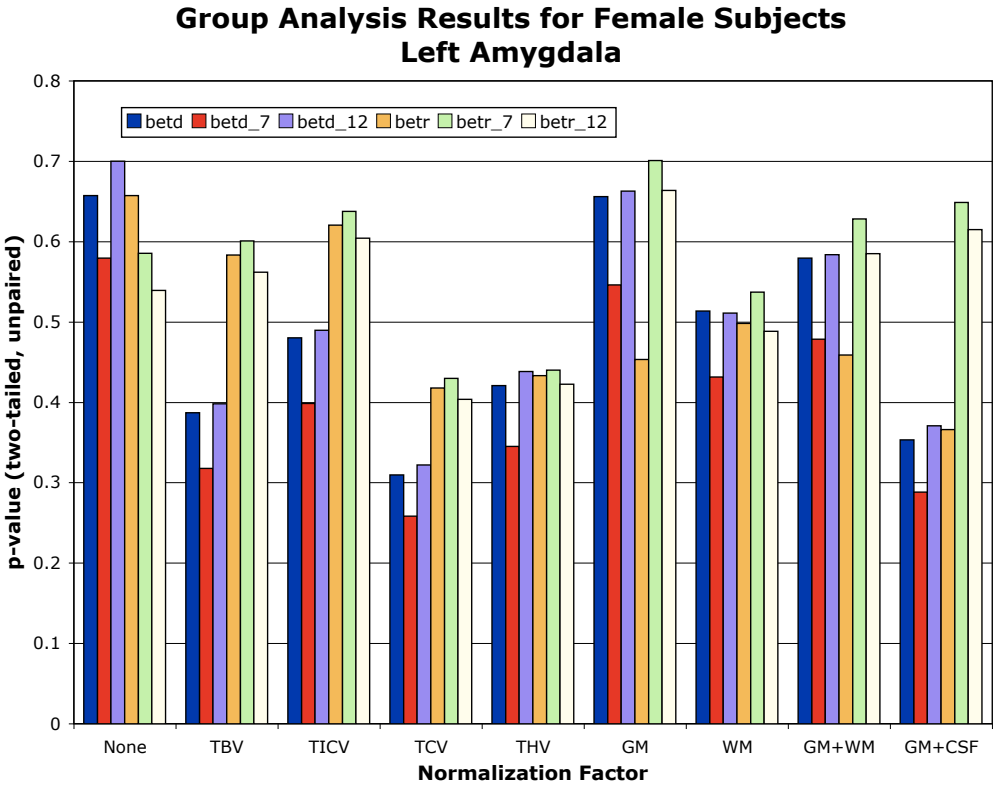

Figure 11: Group analysis results with female subjects for left amygdalar volume corrected by various normalization factors.

3.4. Amygdalar Normalization with ANCOVA

The results of adjusting amygdalar volume with univariate ANCOVA analysis are reported in Supplementary Table 8. With all subjects combined only by gender, the ANCOVA group results resemble those obtained with t-tests for right amygdalar volume (Figure 14). However, significant group differences for left amygdalar volume are a small fraction of those obtained with t-tests (Figure 14). When using gender as an additional fixed factor to examine group by gender interaction effects, there were no significant differences for right or left amygdalar volumes.

Often times when ANCOVA analyses is conducted, t-tests and mean comparisons are used in post hoc analyses to determine the directionality of any observed significant effects. In this study, there were no group by gender interaction effects, which would indicate that further exploration of gender effect need not be investigated. It is clear though, based on the results obtained with t-tests, that there are significant gender effects.

3.5. Repeated Measures ANOVA

Results of the Repeated Measures ANOVA analysis are summarized in Table **??**. The brain segmentation method ($F(1,33) = 10.9$, $\eta^2 = 0.25$, $p = 0.002$) and normalization volume ($F(1,33) = 642.9$, $\eta^2 = 0.95$, $p < 0.001$) proved to be significant sources of variation of amygdalar size for individuals. While the methods for brain segmentation and normalization were significant, the effects varied with normalization volume as a function of group ($F(1,33) = 15.9$, $\eta^2 = 0.39$, $p < 0.005$). Amygdalar sidedness was also a significant source of variation for groups ($F(1,33) = 4.509$, $\eta^2 = 0.12$, $p = 0.04$). Normalization volume had significant interaction effects with brain segmentation method ($F(1,33) = 30.323$, $\eta^2 = 0.49$, $p < 0.001$) and spatial registration ($F(1,33) = 31.7$, $\eta^2 = 0.53$, $p < 0.001$).

4. Conclusions

The motivation for this study was to investigate the effect of various amygdalar volume normalization methods on group analysis results. Three strategies for spatial, size, and tissue normalization were explored. Spatial normalization was conducted with 7- and 12-DOF transformations. Size normalization included measures of total brain volume, total intracranial volume, total cranial volume, total head volume. Tissue normalization was conducted with volume measures of gray matter, white matter, gray matter

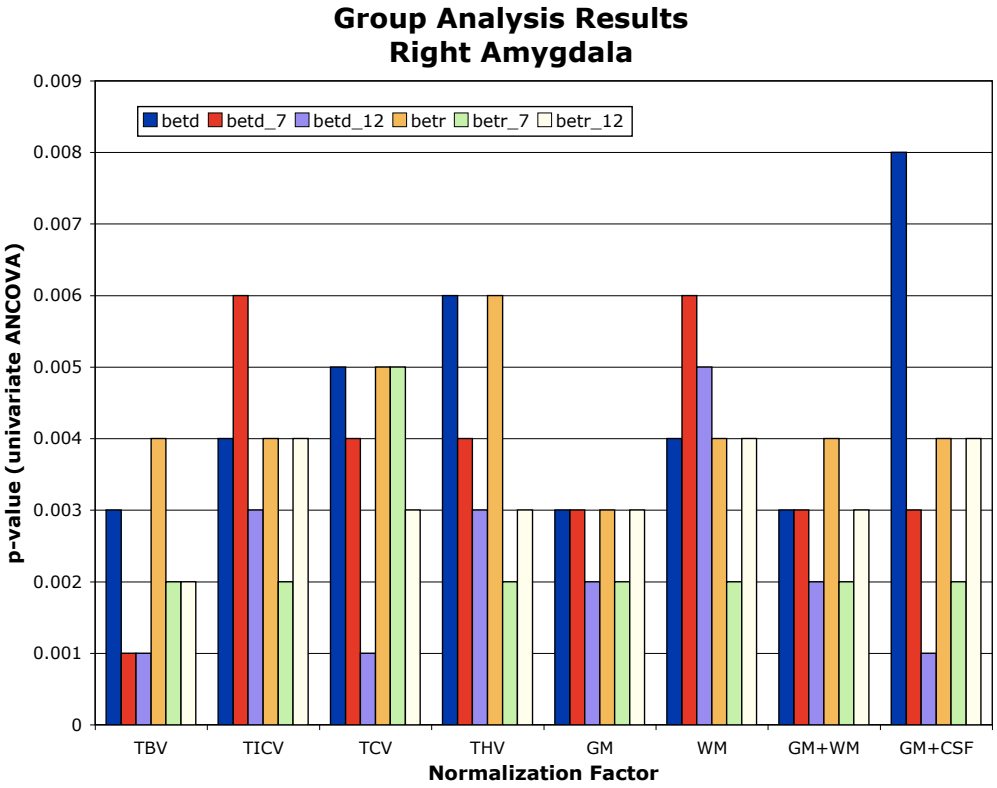

Figure 12: Group analysis results with all subjects combined for right amygdalar volume corrected by various normalization factors.

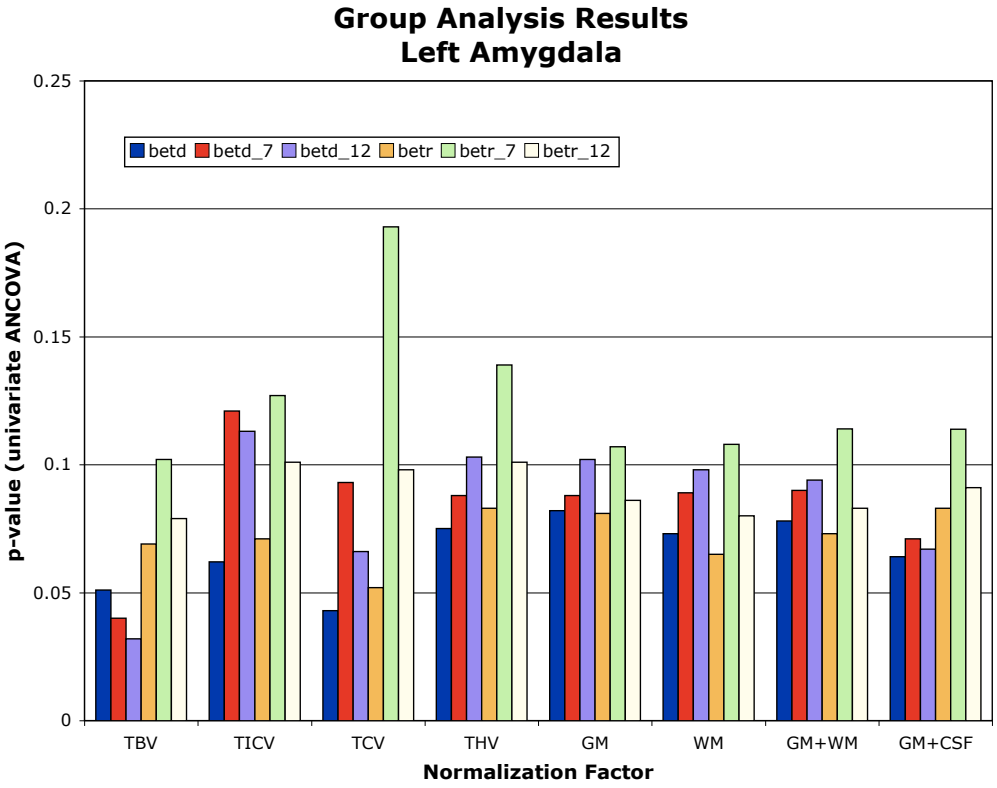

Figure 13: Group analysis results with all subjects combined for left amygdalar volume corrected by various normalization factors.

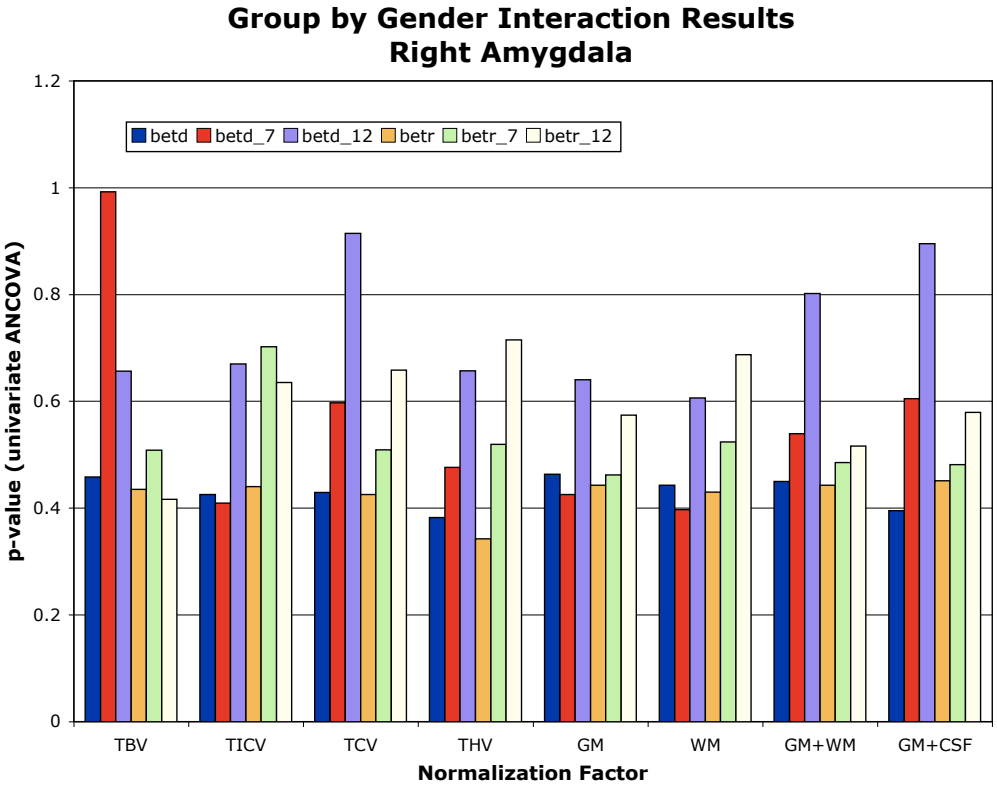

Figure 14: Group by gender interaction effect for right amygdalar volume corrected by various normalization factors.

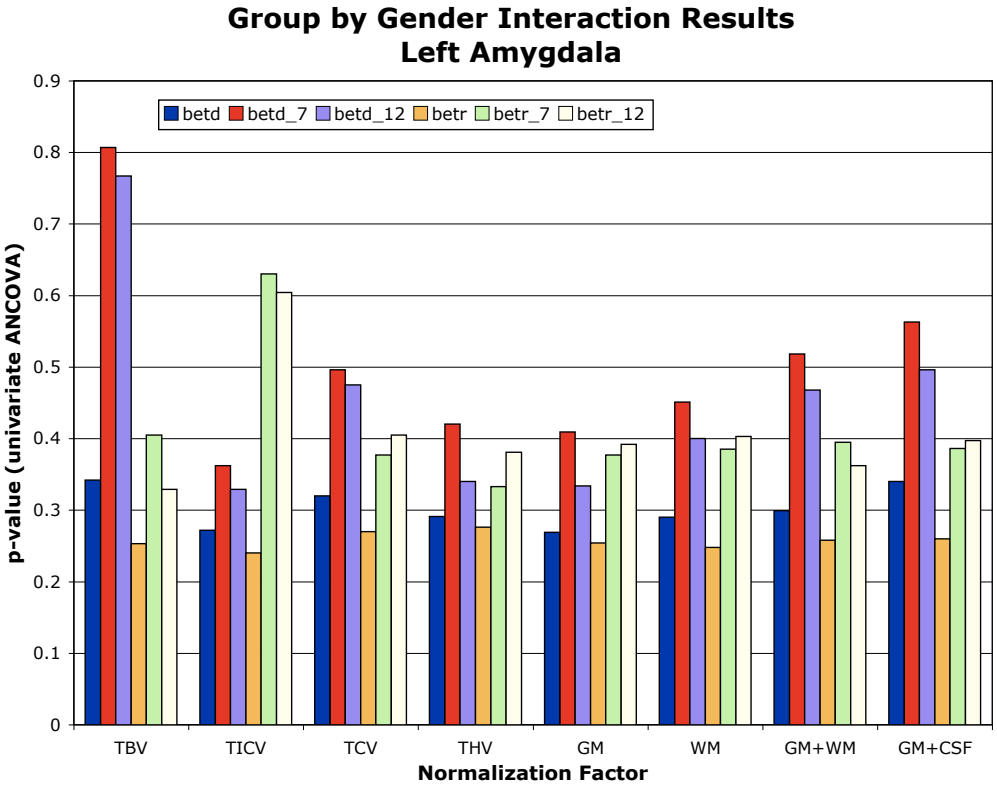

Figure 15: Group by gender interaction effect for left amygdalar volume corrected by various normalization factors.

and white matter, and gray matter and cerebrospinal fluid. Two methods of skull stripping were also explored. Group analysis was performed with unpaired t-tests by dividing amygdalar volume by the various size and tissue volumes, univariate ANCOVA analysis with group and gender as fixed factors and various size and tissue volumes as covariates, and repeated measures ANOVA with four within-subject factors (sidedness, spatial transformation, tissue volume) and one between-subjects factor (group).

Results of this study indicate that group analysis of amygdalar volume are highly dependent on the method of normalization including volumetric normalization factor, skull stripping method, and spatial transformation. We also found that sidedness, group, and gender are significant effects in late-life depression.

We believe that this may partially explain the inconsistent findings reported for amygdalar volumetry in mood disorder research. Although, this research study focused on amygdalar volumetry in the LLD population, results are likely extensible to other volumetric studies. Since there is no standard for volume normalization, investigators should be diligent in reporting the details of their normalization protocol, and be cautious when interpreting and comparing results from published studies.

References

Aging, N.I.o., 2007. Growing Older in America: The Health and Retirement Study. Technical Report. National Institutes of Health.

Altshuler, L.L., Bartzokis, G., Grieder, T., Curran, J., Mintz, J., 1998. Amygdala enlargement in bipolar disorder and hippocampal reduction in schizophrenia: an mri study demonstrating neuroanatomic specificity. Archives of General Psychiatry 55, 663–664.

Beekman, A.T.F., Beurs, E.d., Balkom, A.J.L.M.v., Deeg, D.J.H., Dyck, R.v., Tilburg, W.v., 2000. Anxiety and depression in later life: Co-occurrence and communality of risk factors. American Journal of Psychiatry 157, 89–95.

Bellis, M.D.D., Casey, B.J., Dahl, R.E., Birmaher, B., Williamson, D.E., Thomas, K.M., Axelson, D.A., Frustaci, K., Boring, A.M., Hall, J., Ryan, N.D., 2000. A pilot study of amygdala volumes in pediatric generalized anxiety disorder. Biological Psychiatry 48, 51–57.

Bremner, J.D., Narayan, M., Anderson, E.R., Staib, L.H., Miller, H.L., Charney, D.S., 2000. Hippocampal volume reduction in major depression. American Journal of Psychiatry 157, 115–118.

Bremner, J.D., Randall, P., Scott, T.M., Bronen, R.A., Seibyl, J.P., Southwick, S.M., Delaney, R.C., McCarthy, G., Charney, D.S., Innis, R.B., 1995. Mri-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. American Journal of Psychiatry 152, 973–981.

Butters, M.A., Young, J.B., Lopez, O.L., Aizenstein, H.J., Mulsant, B.H., Reynolds, C.F., DeKosky, S.T., Becker, J.T., 2008. Pathways linking late-life depression to persistent cognitive impairment and dementia. Dialogues in Clinical Neuroscience 10, 345–57.

Convit, A., McHugh, P., Wolf, O.T., Leon, M.J.d., Bobinski, M., Santi, S.D., 1999. Mri volume of the amygdala: A reliable method allowing separation from the hippocampal formation. Psychiatry Research 90, 113–123.

Cox, R.W., 1996. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research 29, 162–173.

Fagiolo, G., Waldman, A., Jahnal, J.V., 2008. A simple procedure to improve fmrib software library brain extraction tool performance. The British Journal of Radiology 81, 250–251.

Frodl, T., Meisenzahl, E., Zetzsche, T., Bottlender, R., Born, C., Groll, C., Jager, M., Leinsinger, G., Hahn, K., Moller, H.J., 2002. Enlargement of the amygdala in patients with a first episode of major depression. Biological Psychiatry 51, 708–714.

Giedd, J.N., Vaituzis, A.C., Hamburger, S.D., Lange, N., Rajapakse, J.C., Kaysen, D., 1996. Quantitative mri of the temporal lobe, amygdala, and hippocampus in normal human development: ages 4-18. Journal of Comparitive Neurology 366, 223–230.

Gunten, A.v., Fox, N.C., Cipolotti, L., Ron, M.A., 2000. A volumetric study of hippocampus and amygdala in depressed patients with subjective memory problems. Journal of Neuropsychiatry and Clinical Neurosciences 12, 493–498.

Hastings, R.S., Parsey, R.V., Oquendo, M.A., Arango, V., Mann, J.J., 2004. Volumetric analysis of the prefrontal cortex, amygdala, and hippocampus in major depression. Neuropsychopharmacology 29, 952–959.

Health, N.I.o., 2007. About depression. http://nihseniorhealth.gov .

Hensel, A., Wolfe, H., Dieterlen, T., Riedel-Heller, S., Arendt, T., Gertz, H.J., 2005. Morphometry of the amygdala in patients with questionable dementia and mild dementia. Journal of the Neurological Sciences 238, 71–74.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825–841.

Jenkinson, M., Pechaud, M., Smith, S., 2005. Bet2: Mr-based estimation of brain, skull and scalp surfaces, in: Eleventh Annual Meeting of the Organization for Human Brain Mapping.

Kates, W.R., Abrams, M.T., Kaufmann, W.E., Breiter, S.N., Reiss, A.L., 1997. Reliability and validity of mri measurement of the amygdala and hippocampus in children with fragile x syndrom. Psychiatry Research 75, 31–48.

Keller, J., Shen, L., Gomez, R.G., Garrett, A., Solvason, H.B., Reiss, A., Schatzberg, A.F., 2008. Hippocampal and amygdalar volumes in psychotic and nonpsychotic unipolar depression. American Journal of Psychiatry AiA, 1–9.

Kluver, H., Bucy, P.C., 1939. Preliminary analysis of the temporal lobes in monkeys. Archives of Neurological Psychiatry 42, 979–1000.

Lange, C., Irle, E., 2004. Enlarged amygdala volume and reduced hippocampal volume in young women with major depression. Psychological Medicine 34, 1059–1064.

McCarley, R.W., Shenton, M.E., 2008. MR image acquisition and image processing tools and neuroanatomical regions of interest. Technical Report. Harvard University.

Mervaala, E., Fohr, J., Kononen, M., Valkonen-Korhonen, M., Vainio, P., Partanen, K., Partanen, J., Tiihonen, J., Viinamaki, H., Karjalainen, A.K., Lehtonen, J., 2000. Quantitative mri of the hippocampus and amygdala in severe depression. Psychological Medicine 30, 117–125.

Munn, M.A., Alexopoulos, J., Nishino, T., Babb, C.M., Flake, L.A., Singer, T., Ratnanather, J.T., Huang, H., Todd, R.D., Miller, M.I., Botteron, K.N., 2007. Amygdala volume analysis in female twins with major depression. Biological Psychiatry 62, 415–422.

Pruessner, J.C., Li, L.M., Series, W., Pruessner, M., Colins, D.L., Kabani, N., Lupien, S., Evans, A.C., 2000. Volumetry of hippocampus and amygdala with high-resolution mri and three-dimensional analysis software : minimizing the discrepancies between laboratories. Cerebral Cortex 10, 433–442.

Rosso, I.M., Cintron, C.M., Steingard, R.J., Renshaw, P.F., Young, A.D., Yurgelun-Todd, D.A., 2005. Amygdala and hippocampus volumes in pediatric major depression. Biological Psychiatry 57, 21–26.

Sheline, Y.I., Gado, M.H., Price, J.L., 1998. Amygdala core nuclei volumes are decreased in recurrent major depression. NeuroReport 9, 2023–2028.

Siegle, G.J., Steinhauer, S.R., Thase, M.E., Stenger, A.V., Carter, C.S., 2002. Can't shake that feeling: Event-related fmri assessment of sustained amygdala activity in response to emotional information in depressed individuals. Biological Psychiatry 51, 693–707.

Smith, S.M., 2002. Fast robust automated brain extraction. Human Brain Mapping 17, 143–155.

Tamburo, R.J., Siegle, G.J., Stetten, G.D., Cois, C.A., Butters, M.A., III, C.F.R., Aizenstein, H.J., 2008. Amygdalae morphometry in late-life depression. International Journal of Geriatric Psychiatry .

Velakoulis, D., Wood, S.J., Wong, M.T.H., McGorry, P.D., Yung, A., Phillips, L., Smith, D., Brewer, W., Proffitt, T., Ddesmond, P., Pantelis, C., 2006. Hippocampal and amygdala volumes according to psychosis stage and diagnosis. Archives of General Psychiatry 63, 139–149.

Watson, C., Andermann, F., Gloor, P., Jones-Gotman, M., Peters, T., Evans, A., 1992. Anatomic basis of amygdaloid and hippocampal volume measurement by magnetic resonance imaging. Neurology 42, 1743–1750.

Xia, J., Chen, J., Zhou, Y., Zhang, J., Yang, B., Xia, L., Wang, C., 2004. Volumetric mri analysis of the amygdala and hippocampus in subjects with major depression. J. Huazhong Univ. Sci. Technol. Med. Sci. 24, 500–502, 506.

Yoo, T. (Ed.), 2004. Insight Into Images. A. K. Peters, Wellesley.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain mr images through a hidden markov random field model and the expectation maximization algorithm. IEEE Trans. on Medical Imaging 20, 45–57.

5. Supplementary Data

Table 1: Amygdalar volumes with and without the application of a spatial transformation. Brains segmented with the BET-default and BET-recursive methods. Data presented as mean volume followed by standard deviation in parentheses. Significance at $p < 0.05$ is shown in bold.

Method	**Group (N)**	**Right Amygdalar Volume (mm^3)**	**Left Amygdalar Volume (mm^3)**
Absolute Volume	Control (19)	1841.4 (384.6)	1839.7 (379.0)
	LLD (16)	1391.5 (390.9)	1542.8 (488.5)
	p-value:	**0.0018**	0.058
	Control, Male (11)	1919.4 (427.4)	1933.4 (397.7)
	LLD, Male (9)	1377.9 (427.2)	1485.8 (534.6)
	p-value	**0.012**	**0.02**
	Control, Female (8)	1734.1 (311)	1710.8 (332.8)
	LLD, Female (7)	1409 (371.6)	1616.1 (452.1)
	p-value:	0.094	0.66
BET-default, 7-DOF	Control (19)	1961.0 (404.4)	1943.8 (444.7)
	LLD (16)	1502.4 (458.8)	1635.9 (524.9)
	p-value:	**0.004**	0.074
	Control, Male (11)	1956.3 (415.4)	1907.2 (469.6)
	LLD, Male (9)	1374.44 (414.9)	1467.8 (490.8)
	p-value	**0.006**	0.058
	Control, Female (8)	1967.6 (417.1)	1994.2 (434.1)
	LLD, Female (7)	1666.8 (490.4)	1852.1 (520.1)
	p-value:	0.23	0.58
BET-default, 12-DOF	Control (19)	1965.7 (414.8)	1937.8 (428)
	LLD (16)	1489.8 (470.5)	1647.2 (535.4)
	p-value:	**0.004**	0.09
	Control, Male (11)	1949.1 (441.8)	1902.4 (460)
	LLD, Male (9)	1385.4 (456.4)	1460.4 (475.8)
	p-value:	**0.013**	0.051
	Control, Female (8)	1988.5 (403.4)	1986.4 (405.7)
	LLD, Female (7)	1623.8 (488.4)	1887.3 (543.4)
	p-value:	0.14	0.7
BET-recursive, 7-DOF	Control (19)	1935.9 (394.1)	1875 (401.2)
	LLD (16)	1442.7 (410)	1598.2 (493.3)
	p-value:	**0.001**	0.082
	Control, Male (11)	1934.7 (428)	1855.73 (452.6)
	LLD, Male (9)	1358.44 (407.2)	1457.7 (484.5)
	p-value:	**0.0066**	**0.023**
	Control, Female (8)	1938 (371.1)	1902 (346.2)
	LLD, Female (7)	1551 (418.1)	1779 (477.2)
	p-value:	0.084	0.58
BET-recursive, 12-DOF	Control (19)	1928.8 (432.2)	1894.4 (409.1)
	LLD (16)	1451.2 (408.8)	1605 (475.8)
	p-value:	**0.002**	0.066
	Control, Male (11)	1919.6 (460.7)	1858.6 (430.4)
	LLD, Male (9)	1384.2 (421.1)	1455.6 (437.4)
	p-value:	**0.014**	**0.016**
	Control, Female (8)	1941.6 (420.4)	1943.5 (401.2)
	LLD, Female (7)	1537.3 (407.45)	1797.14 (484)
	p-value:	0.082	0.54

Table 2: Group analysis with volume of segmentations using BET-default, and with and without spatial normalization. Data presented as mean (standard deviation) volume in cm^3. p-values were calculated between groups by diagnosis and gender.

Method	**Group**	**TBV**	**TICV**	**TCV**	**THV**
BET-default 0-DOF	Control	1389.3 (156.4)	1456.2 (161.1)	1806.9 (189.5)	3065.0 (257.3)
	LLD	1404.3 (143.8)	1479.6 (154.9)	1874.9 (178.0)	3135.8 (227.7)
	p-value:	0.77	0.66	0.28	0.39
	Male Control	1472.7 (135.8)	1539.1 (151.8)	1912.1 (170.8)	3218.5 (209.8)
	Male LLD	1460.9 (149.4)	1552.6 (152.7)	1950.1 (180.5)	3237.8 (228.6)
	p-value:	0.86	0.85	0.64	0.85
	Female Control	1274.7 (103.1)	1342.2 (90.1)	1662.2 (96.4)	2853.9 (139.6)
	Female LLD	1331.6 (105)	1385.7 (103)	1778.2 (128.1)	3004.7 (155.8)
	p-value:	0.31	0.40	0.076	0.073
BET-default 7-DOF	Control	1457.7 (45.5)	1528.8 (60.7)	1897.5 (69.8)	3194.6 (191.1)
	LLD	1481.8 (62.8)	1560.6 (55.6)	1979.7 (92.6)	3290.5 (233.1)
	p-value:	0.21	0.12	**0.007**	0.2
	Male Control	1463.4(46.9)	1528.3 (45.2)	1899.9 (55.3)	3183.5 (142.4)
	Male LLD	1449.7 (27.1)	1541.2 (27.4)	1936.9 (54.5)	3206.0 (169.2)
	p-value:	0.43	0.44	0.15	0.76
	Female Control	1449.9 (45.4)	1529.4 (80.9)	1894.2 (90.2)	3209.9 (52.7)
	Female LLD	1523.1 (73.2)	1585.5 (73.8)	2034.8 (105.8)	3399.0 (42.6)
	p-value:	**0.046**	0.18	**0.018**	0.19
BET-default 12-DOF	Control	1459.0 (46.4)	1529.8 (59.3)	1898.9 (72.4)	3191.9 (196.1)
	LLD	1488.2 (69.9)	1567.0 (62.3)	1988.0 (99.3)	3294.5 (226.5)
	p-value:	0.17	0.08	**0.006**	0.17
	Male Control	1466.6 (46.6)	1531.4 (43.5)	1903.3 (58.2)	3183.2 (152.3)
	Male LLD	1458.2 (37.7)	1549.8 (32.0)	1947.7 (57.6)	3219.8 (170.0)
	p-value:	0.66	0.29	0.11	0.62
	Female Control	1448.6 (47.1)	1527.6 (79.5)	1892.7 (92.5)	3203.8 (53.9)
	Female LLD	1526.7 (85.1)	1589.2 (85.6)	2039.9 (121.2)	3390.6 (265.6)
	p-value:	0.059	0.18	**0.024**	0.19

Table 3: Group analysis with volume of segmentations using BET-recursive, and with and without spatial normalization. Data presented as mean (standard deviation) volume in cm^3. p-values were calculated between groups by diagnosis and gender.

Method	**Group**	**TBV**	**TICV**	**TCV**	**THV**
BET-recursive 0-DOF	Control	1369.8 (146.1)	1430.5 (153.4)	1782.1 (181.4)	3053.4 (266.6)
	LLD	1371.1 (145.9)	1443.7 (156.3)	1834.4 (179.5)	3111.1 (222.7)
	p-value:	0.98	0.80	0.40	0.49
	Male Control	1443.6 (132.1)	1505.1 (152.2)	1879.6 (168.7)	3209.1 (224.6)
	Male LLD	1443.7 (142.9)	1527.8 (149)	1924.2 (176.8)	3211.4 (236.2)
	p-value:	0.99	0.74	0.57	0.98
	Female Control	1268.2 (98)	1327.814875 (81.4)	1648.023 (92.2)	2839.367375 (142.2)
	Female LLD	1279.7 (92.5)	1338.6 (92.6)	1721.8 (112.2)	2983.2 (126.2)
	p-value:	0.82	0.82	0.19	0.058
BET-recursive 7-DOF	Control	1417.2 (24.4)	1480.7 (48.8)	1845.4 (69.3)	3140.3 (175.2)
	LLD	1417.3 (20.3)	1492.4 (37.9)	1898.2 (52.4)	3206.4 (201.7)
	p-value:	0.99	0.43	**0.015**	0.31
	Male Control	1408.6 (19.3)	1467.7 (36.4)	1834.2 (56.7)	3119.8 (112.2)
	Male LLD	1414.0 (22.9)	1496.2 (39.6)	1885.9 (60.7)	3143.1 (168.3)
	p-value:	0.58	0.12	0.068	0.73
	Female Control	1429.0 (26.8)	1498.6 (60.0)	1860.8 (85.4)	3168.6 (243.7)
	Female LLD	1421.4 (17.2)	1487.6 (38.0)	1914.0 (37.7)	3287.8 (224.0)
	p-value:	0.52	0.67	0.14	0.34
BET-recursive 12-DOF	Control	1415.6 (25.8)	1478.9 (45.6)	1843.5 (71.7)	3133.9 (175.2)
	LLD	1416.4 (23.8)	1491.4 (38.4)	1897.0 (52.0)	3198.0 (196.8)
	p-value:	0.93	0.38	**0.016**	0.32
	Male Control	1408.8 (21.8)	1467.8 (35.8)	1834.9 (64.4)	3115.7 (124.5)
	Male LLD	1416.3 (26.6)	1498.4 (37.6)	1888.8 (56.4)	3144.5 (166.4)
	p-value:	0.51	0.082	0.062	0.67
	Female Control	1425.0 (29.4)	1494.0 (55.3)	1855.2 (83.8)	3159.0 (235.7)
	Female LLD	1416.5 (21.7)	1482.5 (40.5)	1907.5 (47.9)	3266.7 (223.9)
	p-value:	0.53	0.65	0.16	0.38

Table 4: Group analysis with tissue segmentations with BET-default and spatial transformations. Data presented as mean (standard deviation) volume in cm^3. p-values were calculated between groups by diagnosis and gender.

Method	**Group**	**GM**	**WM**	**GM+WM**	**GM+CSF**
BET-default 0-DOF	Control	614.3 (61.6)	564.6 (95.0)	1178.9 (141.8)	958.7 (83.7)
	LLD	610.3 (69.2)	558.7 (85.1)	1169.0 (133.0)	982.0 (87.4)
	p-value:	0.86	0.85	0.83	0.43
	Control, Male	575.2 (52.2)	555.3 (84.2)	1130.5 (124.8)	910.9 (59.3)
	Control, Male	567.0 (70.8)	542.0 (85.0)	1109.2 (137.6)	914.9 (74.0)
	p-value:	0.78	0.73	0.72	0.90
	Control, Female	535.5 (55.8)	455.5 (49.5)	991.0 (87.3)	817.3 (63.8)
	Control, Female	539.2 (52.0)	463.9 (36.0)	1003.1 (62.7)	864.1 (82.4)
	p-value:	0.89	0.71	0.76	0.25
BET-default 7-DOF	Control	587.8 (46.3)	536.1 (40.4)	1123.9 (49.4)	917.1 (44.6)
	LLD	587.6 (64.4)	534.0 (35.9)	1121.7 (68.7)	943.8 (62.6)
	p-value:	0.99	0.87	0.91	0.16
	Control, Male	572.3 (42.6)	549.5 (39.4)	1121.8 (53.4)	907.6 (51.4)
	LLD, Male	564.0 (56.7)	535.8 (35.6)	1099.8 (58.2)	910.2 (42.6)
	p-value:	0.72	0.43	0.39	0.9
	Control, Female	609.2 (45.1)	517.6 (36.1)	1126.8 (46.6)	930.1 (31.7)
	LLD, Female	618.1 (64.5)	531.8 (39.1)	1149.8 (75.1)	987.1 (59.1)
	p-value:	0.77	0.48	0.50	0.049
BET-default 12-DOF	Control	588.7 (48.0)	536.5 (39.4)	1125.2 (50.3)	918.0 (47.2)
	LLD	590.3 (66.7)	536.2 (36.9)	1126.4 (73.0)	948.0 (66.9)
	p-value:	0.94	0.98	0.95	0.14
	Control, Male	573.9 (45.4)	550.8 (36.4)	1124.7 (53.8)	909.6 (55.2)
	LLD, Male	567.0 (57.9)	538.9 (35.9)	1105.9 (60.3)	915.5 (47.7)
	p-value:	0.77	0.47	0.48	0.80
	Control, Female	609.0 (46.5)	516.9 (36.6)	1125.9 (48.7)	929.6 (33.5)
	LLD, Female	620.2 (69.2)	532.6 (40.7)	1152.9 (83.8)	989.9 (67.3)
	p-value:	0.72	0.45	0.47	0.062

Table 5: Group analysis with tissue segmentations with BET-recursive and spatial transformations. Data presented as mean (standard deviation) volume in cm^3. p-values were calculated between groups by diagnosis and gender.

Method	Group	GM	WM	GM+WM	GM+CSF
BET-recursive 0-DOF	Control	646.0 (59.4)	483.8 (74.7)	1211.0 (124.9)	881.7 (78.8)
	LLD	642.1 (73.7)	488.4 (73.0)	1130.5 (122.4)	879.5 (90.5)
	p-value:	0.87	0.86	0.92	0.94
	Control, Male	679.2 (66.2)	531.8 (75.9)	1190.1 (56.0)	934.5 (86.1)
	LLD, Male	667.4 (91.5)	537.8 (72.0)	1205.2 (126.1)	919.4 (111.2)
	p-value:	0.75	0.86	0.75	0.74
	Control, Female	612.9 (51.6)	441.0 (56.9)	1053.9 (84.8)	831.9 (70.2)
	LLD, Female	640.9 (55.2)	450.6 (42.9)	1091.5 (77.2)	877.3 (81.8)
	p-value:	0.33	0.71	0.38	0.27
BET-recursive 7-DOF	Control	670.0 (38.6)	498.8 (31.6)	1168.8 (31.0)	914.2 (43.5)
	LLD	665.2 (54.0)	503.6 (36.2)	1168.8 (31.0)	910.3 (47.8)
	p-value:	0.77	0.68	0.99	0.80
	Control, Male	654.5 (24.4)	504.9 (31.0)	1159.4 (30.5)	897.7 (31.0)
	LLD, Male	652.0 (61.1)	511.7 (39.9)	1163.7 (32.9)	898.7 (54.1)
	p-value:	0.91	0.68	0.76	0.96
	Control, Female	691.4 (45.7)	490.3 (32.5)	1181.8 (28.3)	936.9 (49.7)
	LLD, Female	682.2 (41.2)	493.2 (30.3)	1175.4 (29.5)	925.3 (36.5)
	p-value:	0.69	0.86	0.68	0.61
BET-recursive 12-DOF	Control	669.9 (40.8)	497.8 (29.4)	1167.7 (31.4)	913.7 (46.7)
	LLD	665.0 (55.5)	503.3 (35.1)	1168.2 (34.3)	909.8 (49.6)
	p-value:	0.77	0.62	0.96	0.81
	Control, Male	654.9 (28.7)	504.6 (27.9)	1159.5 (31.0)	898.3 (37.4)
	LLD, Male	653.3 (62.1)	512.6 (38.1)	1165.9 (35.0)	900.1 (56.2)
	p-value:	0.94	0.61	0.68	0.93
	Control, Female	690.5 (47.7)	488.4 (30.6)	1178.9 (30.2)	934.9 (52.3)
	LLD, Female	679.9 (45.8)	491.4 (29.1)	1171.3 (36.0)	922.3 (40.2)
	p-value:	0.67	0.85	0.67	0.61

Table 6: Statistical results (two-tailed t-test, p-values) for right amygdalae normalization by various volume measures, two brain segmentation methods, and spatial transformations.

Method	**Group**	**TBV**	**TICV**	**TCV**	**THV**	**GM**	**WM**	**GM+WM**	**GM+CSF**
BET-default, 0-DOF	All	0.0012	0.0017	0.0007	0.0012	0.00078	0.004	0.0017	0.0007
	Male	0.014	0.014	0.011	0.017	0.0078	0.027	0.013	0.01
	Female	0.045	0.063	0.033	0.043	0.092	0.072	0.07	0.041
BET-default, 7-DOF	All	0.0012	0.0018	0.0007	0.0012	0.0011	0.0045	0.0016	0.00067
	Male	0.0063	0.0066	0.005	0.0077	0.0026	0.014	0.0054	0.0042
	Female	0.093	0.12	0.071	0.086	0.16	0.14	0.13	0.086
BET-default, 12-DOF	All	0.00009	0.0014	0.0005	0.0011	0.0009	0.0034	0.0012	0.00054
	Male	0.012	0.012	0.0097	0.015	0.0065	0.021	0.01	0.009
	Female	0.045	0.064	0.032	0.049	0.082	0.077	0.068	0.038
BET-recursive, 0-DOF	All	0.002	0.0025	0.001	0.0016	0.0016	0.0039	0.0019	0.0017
	Male	0.014	0.013	0.01	0.02	0.014	0.018	0.012	0.017
	Female	0.046	0.1	0.054	0.046	0.048	0.077	0.049	0.044
BET-recursive, 7-DOF	All	0.00087	0.0011	0.00043	0.00068	0.00065	0.0019	0.00081	0.00074
	Male	0.0048	0.0057	0.0044	0.0093	0.0026	0.011	0.0046	0.0031
	Female	0.094	0.1	0.056	0.045	0.12	0.088	0.091	0.11
BET-recursive, 12-DOF	All	0.0014	0.0017	0.00073	0.0012	0.0012	0.0028	0.0013	0.0013
	Male	0.01	0.011	0.0092	0.018	0.0063	0.02	0.01	0.0071
	Female	0.083	0.088	0.048	0.041	0.1	0.074	0.078	0.1

Table 7: Statistical results (two-tailed t-test, p-values) for left amygdalae normalization by various volume measures, two brain segmentation methods, and spatial transformations.

Method	**Group**	**TBV**	**TICV**	**TCV**	**THV**	**GM**	**WM**	**GM+WM**	**GM+CSF**
BET-default, 0-DOF	All	0.044	0.037	0.018	0.041	0.064	0.055	0.051	0.024
	Male	0.038	0.036	0.030	0.049	0.042	0.042	0.036	0.039
	Female	0.39	0.48	0.31	0.42	0.66	0.51	0.58	0.35
BET-default, 7-DOF	All	0.038	0.046	0.024	0.043	0.072	0.067	0.061	0.002
	Male	0.066	0.061	0.052	0.073	0.07	0.074	0.065	0.064
	Female	0.32	0.4	0.26	0.34	0.54	0.43	0.48	0.29
BET-default, 12-DOF	All	0.045	0.053	0.028	0.052	0.092	0.08	0.074	0.037
	Male	0.056	0.052	0.043	0.063	0.063	0.063	0.056	0.054
	Female	0.4	0.49	0.32	0.44	0.66	0.51	0.58	0.37
BET-recursive, 0-DOF	All	0.044	0.048	0.024	0.041	0.058	0.048	0.045	0.049
	Male	0.036	0.032	0.027	0.053	0.041	0.04	0.032	0.045
	Female	0.39	0.62	0.42	0.43	0.45	0.5	0.46	0.36
BET-recursive, 7-DOF	All	0.075	0.08	0.044	0.068	0.1	0.076	0.078	0.084
	Male	0.066	0.068	0.06	0.097	0.74	0.076	0.065	0.069
	Female	0.6	0.64	0.43	0.44	0.7	0.54	0.63	0.65
BET-recursive, 12-DOF	All	0.06	0.065	0.034	0.054	0.081	0.06	0.062	0.068
	Male	0.045	0.046	0.04	0.07	0.047	0.056	0.043	0.045
	Female	0.56	0.6	0.4	0.42	0.66	0.49	0.58	0.61

Table 8: Results of univariate ANCOVA analyses with group and gender as independent factors and the various volume measures as covariates. The top/bottom shows results with right/left amygdalar volume as the dependent variable. Reported are p-values for Group and Group by Gender (G *G) interaction effects.

Dependent Variable:	Right Amygdalar	Volume							
Method	**Predictor**	**TBV**	**TICV**	**TCV**	**THV**	**GM**	**WM**	**GM+WM**	**GM+CSF**
BET-default 0-DOF	Group	0.003	0.004	0.005	0.006	0.003	0.004	0.003	0.008
	G * G	0.458	0.425	0.429	0.382	0.463	0.443	0.45	0.395
BET-default 7-DOF	Group	0.001	0.006	0.004	0.004	0.003	0.006	0.003	0.003
	G * G	0.992	0.409	0.597	0.476	0.425	0.397	0.539	0.605
BET-default 12-DOF	Group	0.001	0.003	0.001	0.003	0.002	0.005	0.002	0.001
	G * G	0.656	0.67	0.914	0.657	0.64	0.606	0.802	0.895
BET-recursive 0-DOF	Group	0.004	0.004	0.005	0.006	0.003	0.004	0.004	0.004
	G * G	0.435	0.44	0.425	0.342	0.443	0.43	0.443	0.451
BET-recursive 7-DOF	Group	0.002	0.002	0.005	0.002	0.002	0.002	0.002	0.002
	G * G	0.508	0.702	0.509	0.519	0.462	0.524	0.485	0.481
BET-recursive 12-DOF	Group	0.002	0.004	0.003	0.003	0.003	0.004	0.003	0.004
	G * G	0.416	0.635	0.658	0.715	0.574	0.687	0.516	0.579
Dependent Variable:	Left Amygdalar	Volume							
Method	**Predictor**	**TBV**	**TICV**	**TCV**	**THV**	**GM**	**WM**	**GM+WM**	**GM+CSF**
BET-default 0-DOF	Group	0.051	0.062	0.043	0.075	0.082	0.073	0.078	0.064
	G * G	0.342	0.272	0.32	0.291	0.269	0.29	0.299	0.34
BET-default 7-DOF	Group	0.04	0.121	0.093	0.088	0.088	0.089	0.09	0.071
	G * G	0.807	0.362	0.496	0.42	0.409	0.451	0.518	0.563
BET-default 12-DOF	Group	0.032	0.113	0.066	0.103	0.102	0.098	0.094	0.067
	G * G	0.767	0.329	0.475	0.34	0.334	0.4	0.468	0.496
BET-recursive 0-DOF	Group	0.069	0.071	0.052	0.083	0.081	0.065	0.073	0.083
	G * G	0.253	0.24	0.27	0.276	0.254	0.248	0.258	0.26
BET-recursive 7-DOF	Group	0.102	0.127	0.193	0.139	0.107	0.108	0.114	0.1138
	G * G	0.405	0.63	0.377	0.333	0.377	0.385	0.395	0.386
BET-recursive 12-DOF	Group	0.079	0.101	0.098	0.101	0.086	0.08	0.083	0.091
	G * G	0.329	0.604	0.405	0.381	0.392	0.403	0.362	0.397